Zheng Cui

Nanofabrication

Principles, Capabilities and Limits

Third Edition

Springer

Nanofabrication

Zheng Cui

Nanofabrication

Principles, Capabilities and Limits

Third Edition

Springer

Zheng Cui
Suzhou Institute of Nanotech and Nanobionics
Chinese Academy of Sciences
Suzhou, China

ISBN 978-3-031-62545-9 ISBN 978-3-031-62546-6 (eBook)
https://doi.org/10.1007/978-3-031-62546-6

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2008, 2017, 2024
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

It has been another eight years since the second edition of this book was published, and 16 years since the first edition was published in 2008. I was told by Springer that by end of April 2023 nearly 800 copies of the books have been sold, plus 41,000 downloads from Springer Link by worldwide educational and academic users, which is not bad considering that nanofabrication is a relatively narrow field of engineering discipline. Last May, the editor of Springer invited me to write the third edition. I thought it is the time to update the book, as a lot has happened in nanofabrication technology since the second edition was published.

The most significant development in the last eight years was the revival of Moore's law which has been a law of predicting the trend of integrated circuit (IC) for more than 60 years. I remembered that an article published in *Nature* in February 2016 with the heading of "The chips are down for Moore's law." While it is natural to think that the features dimension of an IC cannot go down forever, which was then predicted that the Moore's law would come to an end by early 2020s, the IC industry came up with a new way to interpret the Moore's law since then. The current "3 nm generation of ICs" (Moore's law) actually need only to have 12 nm half-pitch minimum features to be made (Sect. 11.2 of Chap. 11), which is perfectly possible thanks to the advances in nanofabrication happened in the last 8 years, namely the high-NA EUV lithography (Chap. 2). Now it looks like the Moore's law will go on viable at least for the next 15 years!

Although IC manufacturing represents the apex of nanofabrication, there is only a small group of people who actually get involved in this field because only a handful companies around the world can afford to conduct such hugely expensive business. Most readers of this book may work in diverse fields other than IC manufacturing. Same as the last edition, the new edition maintains the wide coverage of nanofabrication. Most techniques one may encounter in work or in scientific literatures are included, from various top-down lithography and pattern transfer to variety of bottom-up self-assembly approaches. There are also chapters devoted to focused ion beam, nanoimprinting, and scanning probe fabrications. A content unique only to this book is about indirect nanofabrication, which introduces many non-traditional

techniques often called “tricks.” As the book title implies, it is about the principles, capabilities, and limits of these nanofabrication techniques which should be helpful and informative to a wide spectrum of nanofabrication practitioners and novices.

In the last eight years, my own career had a major change again. I have retired since 2020 from my post at the Suzhou Institute of Nanotech and Nanobionics, Chinese Academy of Sciences. Although I have not completely cutoff connections with the Institute, now working as an advisor, one good thing is that I have plenty time to work on the third edition. With the time in hand, I thoroughly read through past and present literatures and carefully reviewed all the chapters in the second edition. The third edition not only has included the most updated information and new knowledges but also has the angles of some views readjusted. Meanwhile, I am now more involved in industry by participating some startup companies, which enables me to have a new perspective on the nanofabrication technologies I am familiar with. All these have helped me to refine the third edition.

Finally, my sincere gratitude goes to the Institute I was working with and all funding bodies which gave me the opportunities, resources, and the platform in the past and present. I am also indebted to my wife, Ling, who put up with me for the time I buried myself working on the book.

Suzhou, China Zheng Cui
March, 2024

Contents

1 Introduction 1
1.1 What Is Nanofabrication? 1
1.2 Classification of Nanofabrication 3
1.3 Purpose of the Book 6

2 Optical Lithography 9
2.1 Introduction 9
2.2 Principle of Optical Projection Lithography 10
2.3 Basics of Photoresists 15
2.3.1 Process of Optical Lithography 15
2.3.2 Characteristics of Photoresists 17
2.4 Optical Lithography at Shorter Wavelengths 21
2.4.1 Deep UV 21
2.4.2 Extreme UV 22
2.4.2.1 EUV Source 23
2.4.2.2 EUV Optics 24
2.4.2.3 EUV Mask 25
2.4.2.4 EUV Resists 26
2.4.3 X-Ray 26
2.5 Optical Lithography at High NA 28
2.6 Immersion Lithography 32
2.7 Optical Lithography at Low k_1 Factor 34
2.7.1 Off-Axis Illumination (OAI) 37
2.7.2 Phase-Shifting Mask (PSM) 38
2.7.3 Optical Proximity Correction (OPC) 44
2.7.4 Photoresists 48
2.7.4.1 Sensitivity 49
2.7.4.2 Contrast 50
2.7.4.3 Line Edge/Linewidth Roughness (LER/LWR) 52

2.7.5 Multi-patterning . . . 55
2.7.6 Design-Technology Co-optimization (DTCO) . . . 57
2.8 Near-Field Optical Lithography . . . 59
2.9 Talbot Optical Lithography . . . 63
2.10 Interferometric Optical Lithography . . . 65
2.11 Maskless Optical Lithography . . . 69
2.12 Two-Photon Polymerization Lithography . . . 72
References . . . 75

3 Electron Beam Lithography . . . 83
3.1 Introduction . . . 83
3.2 Principle of Electron Optics . . . 85
3.2.1 Electron Lens . . . 85
3.2.2 Electron Source . . . 89
3.2.3 Aberrations . . . 91
3.3 Electron Beam Lithography Systems . . . 95
3.3.1 Basic Configuration . . . 95
3.3.2 Key Specifications . . . 97
3.3.3 Vector and Raster Scanning . . . 98
3.3.4 Pattern Fragmentation . . . 99
3.3.5 Commercial E-Beam Lithography Systems . . . 101
3.4 Scattering and Proximity Effect . . . 102
3.4.1 Electron Scattering . . . 103
3.4.2 Proximity Effect and Correction . . . 108
3.4.3 Effect of Secondary Electrons . . . 114
3.4.4 Low Energy E-Beam Lithography . . . 115
3.5 Resist Materials and Processes . . . 119
3.5.1 Sensitivity . . . 119
3.5.2 Contrast . . . 124
3.5.3 Resolution Enhancement Processes . . . 126
3.6 Ultimate Resolution of E-Beam Lithography . . . 130
3.7 High-Throughput E-Beam Lithography . . . 132
3.7.1 Variable Shaped Beam Lithography . . . 133
3.7.2 Multi-E-Beam Lithography . . . 134
References . . . 135

4 Nanofabrication by Focused Ion Beam . . . 141
4.1 Introduction . . . 141
4.2 Ion Sources . . . 142
4.2.1 Liquid Metal Ion Source . . . 142
4.2.2 Plasma Ion Source . . . 145
4.2.3 Gas Field Ion Source . . . 146
4.3 Focused Ion Beam Systems . . . 147
4.4 Ion Scattering in Solid Materials . . . 148
4.5 Ion Sputtering and Deposition . . . 151

4.5.1 Ion Sputtering . . . 151
4.5.2 Ion Beam Assisted Deposition . . . 156
4.6 Focused Ion Beam Direct Nanofabrication . . . 156
4.6.1 Inspecting and Editing Integrated Circuits . . . 157
4.6.2 Repairing Defects of Optical Masks . . . 158
4.6.3 Preparing TEM/STEM Samples . . . 161
4.6.4 Nanostructuring for Scientific Research . . . 162
4.7 Focused Ion Beam Lithography . . . 162
References . . . 165

5 Tip-Based Nanofabrication . . . 169
5.1 Introduction . . . 169
5.2 Principles of Scanning Probe Microscopes . . . 170
5.3 Exposure of Resists . . . 174
5.3.1 Exposure by Electrons . . . 174
5.3.2 Exposure by Photons . . . 177
5.3.3 Exposure by Heat . . . 179
5.4 Surface Modification . . . 180
5.4.1 Hydrogen Depassivation . . . 180
5.4.2 Field-Induced Oxidation . . . 181
5.4.3 Friction-Induced Modification . . . 184
5.4.4 Heat-Induced Modification . . . 187
5.5 Material Deposition . . . 188
5.5.1 Dip-Pen Deposition . . . 188
5.5.2 Field-Induced Deposition . . . 191
5.6 High-Throughput Scanning Probe Lithography . . . 192
5.6.1 High-Speed Scanning . . . 193
5.6.2 Massively Parallel Tip Array . . . 195
References . . . 197

6 Nanoimprint Lithography . . . 201
6.1 Introduction . . . 201
6.2 Thermal Nanoimprint . . . 203
6.2.1 Nanoimprint Stamps . . . 205
6.2.2 Nanoimprint Polymers . . . 207
6.2.3 Demolding . . . 211
6.2.4 Alignment . . . 214
6.3 Room Temperature Nanoimprint . . . 215
6.4 UV Cured Nanoimprint . . . 217
6.4.1 Transparent Stamps . . . 218
6.4.2 UV-Curable Polymers . . . 221
6.4.3 Jet and Flash Imprint Lithography (J-FIL) . . . 225
6.4.4 Substrate Conformal Imprint Lithography (SCIL) . . . 228
6.4.5 Alignment Through Transparent Stamps . . . 230
6.5 Reverse Nanoimprint . . . 232

6.6 Soft Lithography . . . 235
6.6.1 Soft Stamps . . . 236
6.6.2 Microcontact Printing . . . 238
6.6.3 Replication by Capillary Force . . . 241
6.7 Roll-to-Roll Continuous Nanoimprint . . . 245
References . . . 250

7 Nanoscale Pattern Transfer by Etching . . . 257
7.1 Introduction . . . 257
7.2 Wet Chemical Etching . . . 259
7.2.1 Wet Isotropic Etching . . . 259
7.2.2 Wet Anisotropic Etching . . . 261
7.2.3 Metal Assisted Anisotropic Etching . . . 265
7.3 Reactive Ion Etching (RIE) . . . 267
7.3.1 Principle of RIE . . . 267
7.3.2 Process Control in RIE . . . 271
7.3.3 RIE by Inductively Coupled Plasma (ICP) . . . 276
7.4 Deep Reactive Ion Etching (DRIE) . . . 279
7.4.1 Bosch Process . . . 280
7.4.2 Cryogenic Process . . . 283
7.4.3 Critical Issues in DRIE . . . 283
7.4.3.1 Loading Effect . . . 284
7.4.3.2 Microtrenching Effect . . . 285
7.4.3.3 Notching Effect . . . 286
7.4.3.4 Micromasking Effect . . . 288
7.4.3.5 Ion Induced Damage Effect . . . 288
7.5 Atomic Layer Etching (ALE) . . . 289
7.6 Ion Milling . . . 291
References . . . 294

8 Nanoscale Pattern Transfer by Deposition . . . 299
8.1 Introduction . . . 299
8.2 Thin Film Deposition . . . 299
8.3 Pattern Transfer by Liftoff . . . 306
8.4 Pattern Transfer by Plating . . . 313
8.5 Pattern Transfer by Stencil Mask . . . 318
8.6 Pattern Formation by Printing . . . 321
References . . . 324

9 Indirect Nanofabrication . . . 327
9.1 Introduction . . . 327
9.2 Sidewall Lithography . . . 328
9.3 Dimensional Subtraction . . . 332
9.4 Dimensional Addition . . . 337
9.5 Indirect Fabrication of Nanogaps . . . 340
9.6 Super-resolution Patterning . . . 342
References . . . 346

10 Nanofabrication by Self-Assembly 349
10.1 Introduction 349
10.2 Self-Assembly Processes 350
10.3 Self-Assembly of Atoms and Molecules 351
10.3.1 Self-Assembled Monolayer 352
10.3.2 Self-Assembled Supermolecules 354
10.3.3 Evaporation Induced Self-Assembly (EISA) 356
10.3.4 Self-Assembly of Metal-Organic Frameworks (MOFs) 357
10.4 Self-Assembly of Colloidal Particles 358
10.5 Guided Self-Assembly 361
10.5.1 Surface Topography 361
10.5.2 Surface Energy 363
10.5.3 Electrostatic and Magnetic Forces 363
10.5.4 DNA Templates 366
10.6 Nanosphere Lithography 366
10.7 Block Copolymer Lithography 368
10.7.1 Directed Microphase Separation 369
10.7.2 Directed Self-Assembly for IC Manufacturing 371
10.8 Self-Regulated Porous Alumina 373
References 377

11 Applications of Nanofabrication 383
11.1 Introduction 383
11.2 Integrated Circuits 384
11.3 Photonics 391
11.4 Nanotechnology 395
11.4.1 Nanomaterials 396
11.4.2 Nanoelectronics 399
11.4.3 Nanophotonics 400
11.4.4 Nanobiotechnology 401
References 403

Index 407

Chapter 1
Introduction



1.1 What Is Nanofabrication?

Since mankind invented tools, fabrication has been the major production activity throughout human history. The so-called fabrication is to modify raw materials into desired shapes and structures with various tools to acquire certain functionalities. Fabrication is traditionally associated with mechanical processing, i.e., cutting, milling, or molding a work piece into a work part. Many machined work parts are assembled into a complex working system. Some of the conventional mechanical processing can be of nanoscale precision. For example, some of the high precision machining, such as grinding and polishing, can achieve accuracy of a few nanometers. However, these fabrications cannot be called nanofabrication. The fundamental difference between conventional mechanical processing and the nanofabrication described in this book is the dimension of structures or parts to be fabricated. Nanofabrication refers to the fabricated structures and parts themselves are in the dimension of below 100 nanometers. No conventional machine tools can achieve such feats. It is only possible by a fabrication process called planar processing. Jack Kilby, the inventor of integrated circuits and Nobel laureate for his invention, conceived the planar processing method more than 60 years ago, which laid the foundation of modern integrated circuit (IC) manufacturing processes and the associated microfabrication and nanofabrication technologies.

Comparing to conventional mechanical processing, planar processing is a completely different way to make individual parts. There are three distinctive characteristics in the planar processing:

1. Individual parts or structures are formed by planar patterning, instead of conventional direct interaction between a machine tool and raw materials. The planar patterning process is schematically illustrated in Fig. 1.1, which is essentially of two stages: patterning and pattern transfer. A pattern is formed by lithography. In this case it is optical lithography similar to photography whereby the negative is a layer of energy-sensitive polymer called resist coated on a planar supporting

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_1

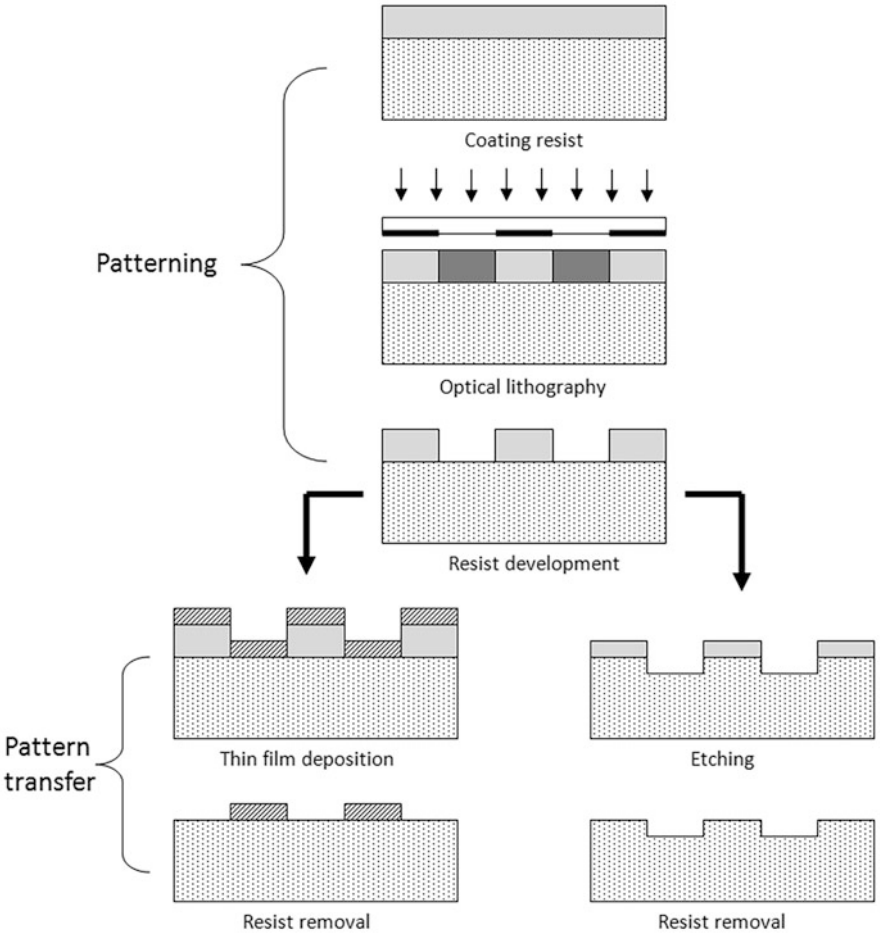


Fig. 1.1 Schematic of planar fabrication process (patterning + pattern transfer)

substrate. The resist layer is exposed and developed, and the developed resist relief pattern is then transferred onto the planar substrate by either etching or additional film deposition. By this way of fabrication, the dimension of a structure is no longer limited by the capability of a machine tool, but by the resolution of a lithography system.

2. Planar process can only make 2D planar structures, not a true 3D object. It at best can make 2.5D structures by superimposition of a series of 2D structures in vertical direction. In contrast to conventional mechanical processing in which individual parts are machined first and a complex system is assembled afterward, planar processing makes a complex system not by assembling individually fabricated parts but by the planar fabrication process itself. The assembly of individual components is realized at different stages of planar processing. As an

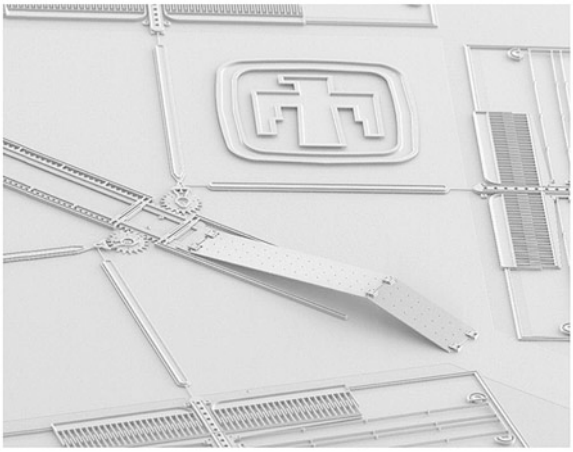


Fig. 1.2 An example of mechanical system fabricated by planar process. (Courtesy Sandia National Laboratories, SUMMiT™ Technologies, www.sandia.gov/mstc)

example, Fig. 1.2 shows a complete mechanical system fabricated by the Sandia National Laboratories of the USA, using planar fabrication processes. It looks like a conventional mechanical system with drive mechanism, which can push a planar structure out of surface plane. What is different from a conventional mechanical system is that the gears are less than 1 mm in diameter, which is beyond the capability of any current precision machining technologies. The individual gears and assembly of drive mechanism were formed in the course of layer by layer planar processing.

3. Planar fabrication is generally a batch process. A batch of devices is made on the same planar substrate through many steps of processes. An 8-inch planar silicon wafer may contain hundreds of individual ICs, such as CPU or memory chips. An example of batch-processed silicon wafer is shown in Fig. 1.3. All the fabrications are done at the wafer level where the hundreds of ICs are made simultaneously with repeated lithography, etching, or deposition processes. Once it is done, individual IC, called "die," is diced and packaged. The batch processing allows mass manufacturing of devices or components.

1.2 Classification of Nanofabrication

The key technology that allows the planar processing to make 2D structures beyond the capability of conventional mechanical machining is the lithography patterning technology. The minimum dimension of feature patterns formed by lithography determines whether a planar processing is microfabrication, i.e., feature dimensions below a millimeter and above 100 nanometers, or nanofabrication, i.e., feature dimension below 100 nanometers. Depending on how a pattern is formed, the lithography can be classified into three types: projection lithography, probe lithography, and mold lithography, as schematically illustrated in Fig. 1.4.

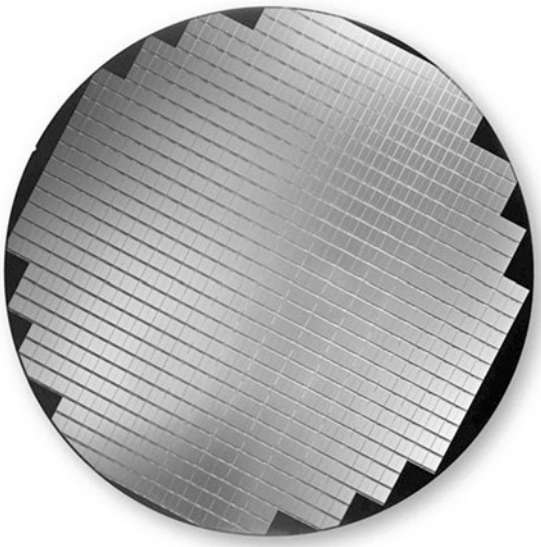


Fig. 1.3 Batch-processed silicon wafer containing hundreds of individual ICs (dies)

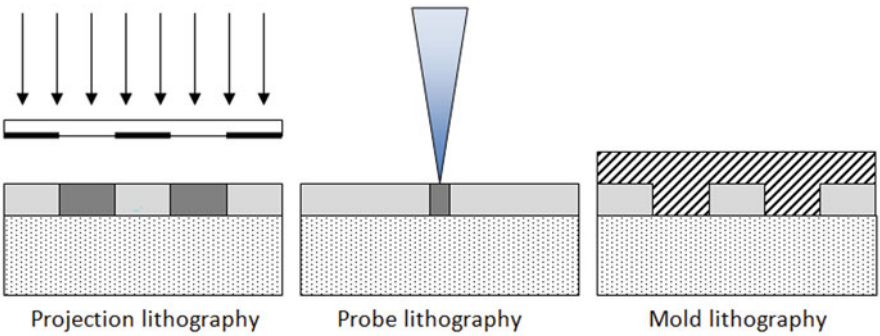


Fig. 1.4 Schematics of three types of lithographic patterning

Projection lithography involves projecting a mask image onto a polymer resist layer coated on a planar substrate. The irradiation that comes through the mask can be a shining light (optical projection lithography), electrons (electron projection lithography), ions (ion projection lithography), and atoms (atom projection lithography). The irradiation causes the exposure of polymer resist which is then developed to become the pattern of mask image for subsequent pattern transfer, as shown in Fig. 1.1. The minimum pattern features that can be exposed depend on the wavelength of irradiation source or the feature size on a mask. Projection lithography is a parallel patterning process, meaning that all the features contained in a mask are imaged simultaneously. High-volume manufacturing of ICs is exclusively done by projection lithography patterning process.

Probe lithography can be considered as the extension of conventional mechanical processing, in which various probes of micro- or nanometer sizes replace the conventional mechanical cutting or milling tools. The probes can be nonsolid probes such as a focused electron beam, an ion beam, or a laser beam. The probes can also be solid types, such as the probes in scanning tunneling microprobes or atomic force microscopes. Nonsolid probes are basically irradiation probes which expose polymer resist to create patterns. Different from the aforementioned projection lithography, exposure of polymer resist performed by an irradiation beam is in sequential manner. The beam probe scans sequentially to create a pattern instead of projecting the whole pattern at once. Therefore, probe lithography is a low-throughput patterning process. Solid probes directly manipulate atoms to create trenches or dents or selectively form oxidation patterns on a substrate material. Alternatively, the solid probes can work like a "fountain pen," transferring liquid to a solid surface to form nanometer-size dots or patterns of only a monolayer of molecules. Nonsolid probes can also directly pattern substrate surface, such as sputtering removal of substrate materials by a focused ion beam or ablation removal of substrate materials by a laser beam, without the help of polymer resist pattern transfer. Whether the fabrication is microscale or nanoscale is mostly dependent on the probe size.

Mold lithography makes patterns by replication of molds. Mold lithography includes nanoimprint, hot embossing, and microinjection molding. Nanoimprint uses master stamps to make nanopatterns by pressing the stamps into a layer of thermally softened or liquid polymer materials. Although both projection and probe fabrications can make micro- and nanostructures as aforementioned, nanoimprint performs the same task at very low cost, because once a stamp is made, it can be used repeatedly to replicate large number of same nanostructures. However, to make the nanoimprint stamps, the projection patterning or probe patterning process has to be used. Nanoimprint has many derivatives, such as thermal nanoimprint, UV-cured nanoimprint, soft contact print, etc. Hot embossing is based on the same principle as the nanoimprint but confined to press a master stamp into thermoplastics only. It is the same technique as for making compact disks. Microinjection molding is for molding of duroplastics or metals. It is an extension of conventional molding process to micrometer range. Both hot embossing and microinjection molding are for replication of microscale structures and objects. Only nanoimprint is qualified as a nanofabrication process.

In addition to the above mentioned planar, probe, and mold patterning technologies, all of which can in one way or another fabricate sub-100 nm structures, there are also self-assembly processes to build nanostructures. These two classes of nanofabrication techniques clearly distinguish each other by their "top-down" and "bottom-up" approaches. In the top-down approach, nanostructures are made by patterning from top. Bottom-up approach relies on self-assembly of nanoscale entities such as nanospheres, nanocrystals, or molecules through spontaneous attraction forces between nanoentities. Bottom-up process can potentially break the feature dimension limit of top-down nanofabrication processes. Self-assembly process generally forms lattice-like structures or periodical arrays. To make useful

nanostructures, self-assembly is normally guided or directed by top-down fabricated template structures.

In general, fabrication of structures larger than 100 nm is considered to be microfabrication. Nanofabrication is referred to fabrication of sub-100 nm structures. Most of nanofabrication technologies, except scanning probes, are developed from microfabrication technologies. Although there is no clear distinction between microfabrication and nanofabrication, from micrometer to nanometer represents not a simple downsizing of structure dimension, but a major advance in fabrication technology. For example, optical lithography used to have its resolution limited by the wavelength of illuminating light. At each wavelength, there is a minimum feature size that can be realized. However, with the advances in optical lithography, its resolution limit has been lowered down continuously in the last several decades. The theoretical limit of optical diffraction to the patterning resolution has already been broken. Now features of less than 10 nm can be fabricated by optical lithography.

1.3 Purpose of the Book

Nanofabrication is a multidisciplinary technology domain covering a wide range of knowledge. Different nanofabrication processes involve different physical or chemical principles. For example, reactive ion etching by plasma discharge involves both physical and chemical processes. Wet etching of materials by acids or alkaline solutions is entirely a chemical process. Although each nanofabrication technology has its theoretical basis in physics or chemistry, it is not the purpose of this book to discuss these theories. The emphasis of this book is to introduce the practical aspects of different nanofabrication technologies. From application point of view, a mature and successful processing technique may not have, or does not need, a thorough theoretical base. In most cases, a successful processing technique is established through many experimental trials and errors. It does not matter if the technique cannot be theoretically explainable, as long as it works successfully to produce the desired results. Basic theoretical knowledge in physics, chemistry, or other branches of science may act as guidance in developing or optimizing a new process technique. While the basic principle of each technology is explained in this book, most of the contents are devoted to practical use of each fabrication technique. It is the author's belief and experience that learning nanofabrication is very much a practical and hands-on process. It is important to know the first principles. The rest is very much down to practicing and mastering the various technologies by using them.

Nanofabrication technologies introduced in this book are essentially about patterning and pattern transfer, as illustrated in Fig. 1.1. Patterning is the most critical step in nanofabrication, which determines how small structures can be made. A nanoscale pattern can be made indirectly by exposure of a polymer layer as a patterned mask. The exposure can be carried out with photons (Chap. 2), electrons (Chap. 3), and ions (Chap. 4). The polymer patterns can also be made without exposure but by scanning a probe (Chap. 5) or imprinting a master template

(Chap. 6). Polymer patterns can be formed even without any tools but by polymer molecules themselves in a process called self-assembly. The typical example is the self-assembly of block copolymers (Chap. 10).

In most cases, a patterned polymer layer itself is not the final product but an intermediate step. A subsequent pattern transfer step is needed to complete the nanofabrication. Pattern transfer is essentially divided into two types, transfer by etching (Chap. 7) which is subtractive process and transfer by deposition (Chap. 8) which is an additive process, as illustrated in Fig. 1.1. Both the pattern transfer processes rely on the polymer pattern as a mask. The final structure is either in the substrate material (etching) or a thin film material (deposition) as the replica of polymer mask pattern. Some tools and techniques can make patterned structures directly without the aid of polymer masks, such as focused ion beam sputtering (Chap. 4) or deposition/removal of materials with scanning probes (Chap. 5). Some nanoscale structures may not be made straightforwardly with a patterning tool but made indirectly to bypass the limitation in resolution for an existing tool (Chap. 9).

There are vast application domains where nanofabrication techniques are pervasively used. This book is not intended to cover all the applications but to select a few most relevant areas where nanofabrication processes described in this book play the key roles. The application areas covered by this book are integrated circuits, photonics, and nanotechnology (Chap. 11).

Though there are a vast number of literatures and published research papers which give details of each nanofabrication technology, they tend to present a positive picture of the technology rather than pointing out what are the shortcomings or telling the negative side of the story. A nanofabrication practitioner knows from his/her own experience that there are no perfect one-suits-all fabrication technologies. Each technology has its own weakness. For example, the optical lithography, especially the latest extreme UV lithography, though is capable of sub-100 nm patterning, is too expensive and not easily accessible; electron beam lithography is of high resolution and high flexibility but low throughput; scanning probes are simple and low-cost systems for sub-100 nm patterning, but the pattern area is small, speed is low, and patterning depth is shallow; nanoimprinting is best suitable for making high density periodical patterns in large quantity; soft lithography is limited in sub-100 nm resolution because of the softness of stamp and lateral diffusion of inks; molecular self-assembly has to be guided with other patterning techniques in order to make any useful pattern structures with reduced defect density. The present book tries to present a balanced view. For each technology introduced, both its advantages and weaknesses are described, so that the readers are made aware of both positive and negative aspects of a nanofabrication technology and aware of its limits.

With the ever expansion of information sources on the Internet, acquiring any knowledge from the Internet has never been so easy, particularly with the help of the latest artificial intelligence (AI). Though the Internet has provided the convenience of finding required information at a stroke of keyboard, it still needs a trained eye to spot the correct and best knowledge among the vast amount of information. Since the book covers so wide topics regarding the nanofabrication, it is written such that the

essential principles of each nanofabrication technique are explained and key knowledge points are described, along with their source of references. Readers can always explore further for more detailed information. The book is intended to be a practical guide and first-hand reference for anyone who wishes to learn about different nanofabrication techniques or to consult with when facing to choose a suitable technique from a "tool box" for a specific application.

Chapter 2
Optical Lithography



2.1 Introduction

Since the planar processing for integrated circuit (IC) manufacturing started more than 60 years ago, optical lithography, or often called photolithography, has become the convenient choice of making planar microstructures. In optical lithography, a mask or photomask, also called reticle, is imaged onto a flat substrate surface coated with a thin layer of polymer material called photoresist. The photon energy is focused into the photoresist, causing polymer chain scission (positive tone photoresist) or cross-linking (negative tone photoresist). The mask pattern is then delineated into the photoresist after development.

In the early days when the feature size of integrated circuits was a few micrometers, optical lithography was done with visible light and was no different from copying photos in a photographic workshop. As the circuit feature dimension shrank, illuminating light with shorter wavelength was employed. This was the time when mercury lamps were used as the illumination source, with UV wavelength emission at g-line ($\lambda = 436$ nm) and i-line ($\lambda = 365$ nm) being the choices of wavelengths for exposure of photoresists. When the circuit dimension shrank to sub-micrometer dimension, short illuminating wavelength became necessary [1]. Deep UV (DUV) at $\lambda = 248$ nm generated from the KrF excimer laser and $\lambda = 193$ nm generated from the ArF excimer laser became the new illumination sources for optical lithography. Many attempts have been tried to explore even shorter wavelengths, such as 157 nm and 126 nm, for optical lithography. They all but failed because there were simply no viable materials to make the refractive lenses to focus them. In the end the extreme UV (EUV) at 13 nm using mirror-based reflective optics became the workhorse for the latest IC manufacturing.

The advances of optical lithography were not only about using short wavelength illumination. Many innovations in the last 20 years have kept optical lithography alive for making generations of IC chips, such as using high numerical apertures (NA), introducing immersion lithography, multiple patterning processes, and a range

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_2

of resolution enhancement techniques (RETs). Optical lithography was able to achieve patterning circuit structures well below the 193 nm illumination wavelength. Now the same innovations are being repeated at the EUV wavelength, which enables the IC manufacturing to enter the sub-3 nm technology node.

Apart from the mainstream development in optical lithography for high-volume IC manufacturing, there is always an army of researchers who are trying to find low-cost alternatives. Interferometric lithography and near-field optical lithography are among the few. These techniques are trying to beat mainstream optical lithography on the cost front but still able to deliver sub-100 nm patterning capability. There is also the maskless optical lithography technique, which promises to deliver sub-100 nm patterning without using the increasingly expensive masks.

Optical lithography has been the subject of many published literatures including previous edition of the book published by the author [2]. The basic issues of optical lithography have been well described in these literatures. The present chapter will not simply repeat those basics but review the photon-based lithography techniques with emphasis on those issues related to producing sub-100 nm structures on a photon-sensitive polymer layer. While the basic principles of various optical lithography techniques are still being mentioned, most of the technical details about the basic optical lithography processes have been omitted. The content has been structured around the central theme of sub-100 nm patterning, the capabilities and limits of current photon-based lithographic techniques, and the ways of overcoming these limits.

2.2 Principle of Optical Projection Lithography

IC manufacturing started in the early days with contact optical lithography, it was soon migrated to projection optical lithography. In contact mode, an optical mask is in direct contact, or having a proximity gap, with a photoresist-coated substrate, and features in the mask are imaged onto the photoresist in 1:1 ratio without reduction. The contact photolithography is performed with a mask aligner and mostly for microscale patterning [3]. Projection photolithography can project a demagnified image onto photoresist (5:1 or 10:1), capable of patterning much smaller features as well as eliminating the mask damage and contamination problems associated with contact photolithography. For submicron patterning, project photolithography prevails.

The basic projection optical lithography system is schematically shown in Fig. 2.1, where the mask is illuminated by a light source (mercury lamp) through a condenser lens and the light passing through the transparent patterns in the mask is focused by a projection lens onto the photoresist layer. A mercury lamp can only generate g-line or i-line wavelength illumination. For deep UV (248 nm or 193 nm wavelength), excimer lasers are used as the light sources. Because of the limited beam size of an excimer laser, the mask is scanned by the laser beam (normally the

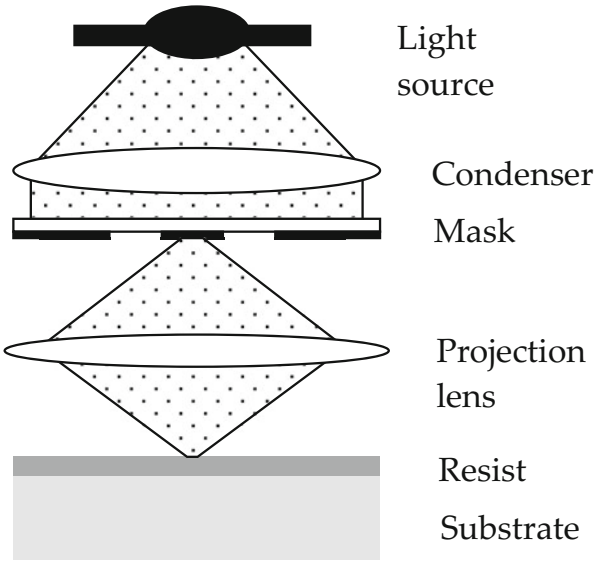


Fig. 2.1 Schematic of projection optical lithography

size of 25 × 33 mm) and projected onto a wafer. For a DUV scanner, the demagnification ratio is 4:1.

The theory of optical imaging in projection photolithography can be very complicated. A comprehensive mathematical description of the theory has been given in a book written by Wong [4]. However, the concept of projection photolithography can be understood easily. A simple conceptual view of diffraction-limited projection photolithography is shown in Fig. 2.2.

A mask feature, which is a transparent opening, can be considered as a slit from which passing light is diffracted in the same way as Young's single slit diffraction. Diffraction is a phenomenon which involves the bending of waves around obstacles. It is described by Huygen's principle that every point on a wave front acts as a source of tiny wavelets that move forward with the same speed as the wave. These wavelets interfere with each other, causing light intensity varies with maxima and minima which are bright and dark fringes, as shown in Fig. 2.3. The angle of each fringe relative to the source can be determined by a simple equation (Eq. 2.1).

$$\sin\varphi = m\frac{\lambda}{a} \quad m = 1, 2, 3, \ldots. \tag{2.1}$$

In Eq. 2.1, ϕ is the diffraction angle of light passing through the slit, λ is the wavelength of light, and a is the width of slit. It is apparent that the spread or expansion of light, represented by the diffraction angle ϕ, is the function of wavelength and the width of slit. For a fixed wavelength, the smaller the slit, the wider the spread of light, as the simulated light intensity distributions shown in Fig. 2.4a for three different widths of slits. The narrower the slit, the wider the distribution. On the other hand, light with shorter wavelength can have narrower distribution of

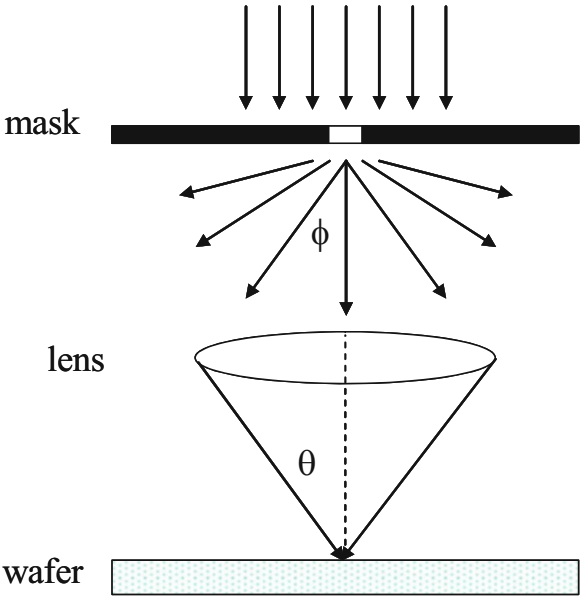


Fig. 2.2 Conceptual view of optical projection imaging

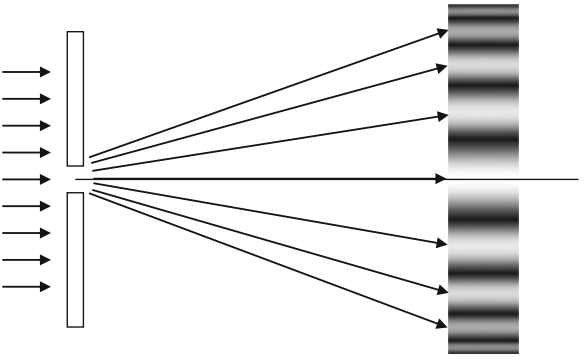


Fig. 2.3 Single slit diffraction

diffraction, as shown in Fig. 2.4b where the light intensity distribution is much narrower for 248 nm wavelength than that for 436 nm wavelength. Referring the slit to a feature opening on a mask, if two of such slits are close to each other, then they may merge into each other, making them irresolvable. Figure 2.5a shows exactly such a case that two light images have merged into one because the slits where the lights come from are too close to each other. They can only be resolved as separate spots providing the two slits are well apart, as shown in Fig. 2.5b.

Light diffraction from a slit, or a mask feature, is only half of the process in a projection optical lithography. Referring to Fig. 2.1, the light coming out of a mask opening has to be focused by a projection lens onto a substrate surface. An optical lens made of glass or quartz is always of limited diameter. The diameter of a lens aperture determines how much or how wide the diffracted light from the mask can be

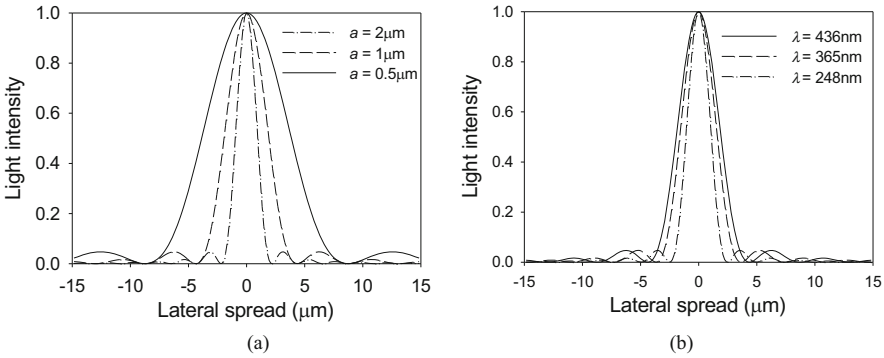


Fig. 2.4 Light intensities of three single slit diffractions (**a**) at 436 nm wavelength, (**b**) for 1 μm slit width

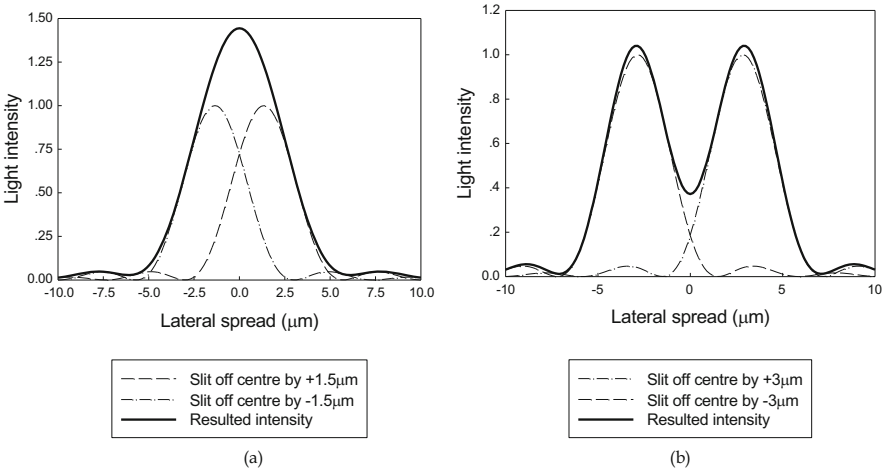


Fig. 2.5 Light intensities of diffraction from separated slits, (**a**) irresolvable image, (**b**) resolvable image

accepted into the lens system. The lens diameter is defined by the numerical aperture (NA) expressed as Eq. 2.2 where n is the refractive index of the medium between the lens and an imaging plane and θ is the half-angle of the maximum cone of light that can enter or exit the lens, or the angle of outmost ray of light entering or exiting the lens as illustrated in Fig. 2.2.

$$\mathrm{NA} = n \sin \theta \tag{2.2}$$

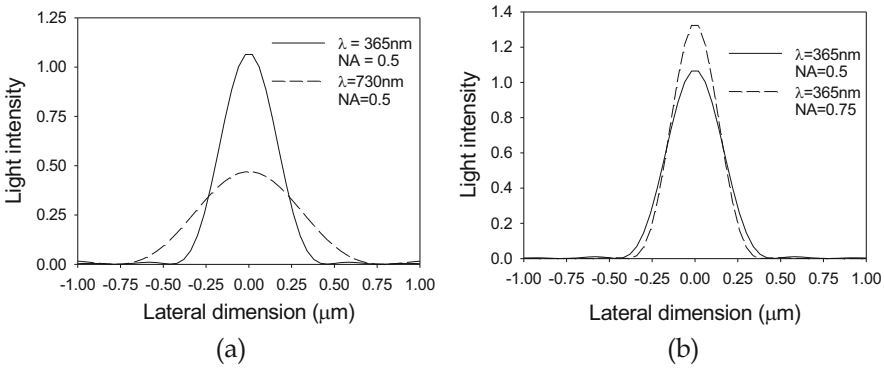


Fig. 2.6 Computer-simulated light intensities imaged at (**a**) different wavelengths and (**b**) different NAs

Skipping the complicated optical imaging theory, a simple equation can define the resolution limit of a projection lithography system, as expressed in Eq. 2.3, where λ is the illumination wavelength, NA is the numerical aperture of imaging optical system, and k_1 is a factor related to a specific imaging process ($k_1 < 1$).

$$R = k_1 \frac{\lambda}{\mathrm{NA}} \tag{2.3}$$

From the above equation, one can intuitively work out that shorter wavelength results in narrow distribution of light diffraction, as proved in Fig. 2.4b. Therefore, more lights can be focused by the lens of fixed diameter (NA). The same applies to larger numerical aperture at a fixed wavelength, i.e., more diffracted lights can be admitted into the lens and be focused onto a substrate if the NA of lens is bigger. Both can result in improved resolution in imaging, because higher-order diffractions (light of large diffraction angle) passing the lens result in more details of the mask feature to be imaged onto the substrate. The effects of wavelength and numerical aperture can be quantitatively evaluated by computer simulation of partially coherent imaging in projection optical lithography. Figure 2.6 shows the simulated light intensity of a mask feature of 0.5 μm linewidth at different illumination wavelengths (Fig. 2.6a) and different NAs (Fig. 2.6b). The simulation was performed by the author's own optical lithography simulation program COMPARE [5].

The k_1 in Eq. 2.3 is a factor that can further enhance the resolution of imaging process by the collective action of many techniques. Sections 2.4, 2.5, 2.6 and 2.7 will describe how the resolution of optical lithography has been improved over the years by using shorter wavelengths, larger numerical apertures, and reduction of the k_1 factor.

2.3 Basics of Photoresists

2.3.1 Process of Optical Lithography

Projection of photomask described in previous section is only one step of optical lithography. The photomask image projected from the optical lens system has to become the surface relief image in photoresist. The whole photoresist imaging process involves 10 steps.

1. Pre-treatment of Substrate
   The substrate surface, which in most cases is a silicon wafer, has to be thoroughly cleaned with chemicals to remove all the stains and particles. Then, it is placed in oven to bake at 150–200 °C for 15 min to half hour to dry completely. Another measure is to coat a thin layer of HMDS (hexamethyldisilazane), which can promote the adhesion between photoresist and silicon surface. Without a good adhesion, photoresist patterns after development may peel off at later stage of process.
2. Coating Photoresist
   Photoresist is coated onto silicon wafer by spin or spray coating process. In a spin coater, photoresist drops at the center of the wafer. By high-speed spinning, the photoresist drops uniformly spreading to the whole surface. The thickness of photoresist can be controlled by different spin speeds and viscosity of photoresist.
3. Pre-exposure Bake
   After coating photoresist, the wafer should be baked. The purpose of pre-exposure bake (pre-bake) is to drive out all the solvent in the photoresist layer so that the photoresist layer becomes completely dry and solid. Pre-bake is also called soft bake, with temperature and baking time specified by photoresist vendors.
4. Optical Exposure
   Silicon wafers coated with photoresist are placed in an optical lithography system for exposure. If it is not the first layer exposure, alignment has to be performed to find the alignment marks and to make sure the overlay of current mask with previously patterned structures on the wafer is accurate. The next step is to choose an appropriate exposure dose. The actual exposure can be done automatically by the stepper. For optical lithography by a mask aligner, a proper contact pressure or proximity gap should be selected before the exposure [3].
5. Post-exposure Bake
   Light wave can be reflected at the interface between photoresist layer and silicon surface during the exposure. Reflected light wave can interfere with the incoming light wave to form standing wave. Such a standing wave can leave its mark in the exposed photoresist. Post-exposure bake (post-bake) can partially eliminate the standing wave effect. However, post-bake may promote lateral diffusion of photoactive compounds in the photoresist, resulting in slight deterioration of photoresist image quality. In recent years, various top and/or bottom

antireflection coatings have been introduced into wafer process. By applying a layer of antireflection coating, the standing wave formation can be effectively prevented.

6. Development

   There are three basic development methods: immersion, spray, and puddle. Immersion development is the simplest, which does not need any special equipment. The exposed wafer is immersed in a developer-filled container for certain time and then taken out to wash off the residual developer. Spray development requires spraying developer onto a spinning wafer. Washing and dry processes are also carried out during wafer spinning. Puddle development combines both the immersion and spray methods. The wafer is first immersed in developer for a certain time and then will undergo spinning for further spray development, cleaning, and drying. Both spray and puddle developments need special equipment.

7. Descum

   Development and washing cannot completely remove the photoresist. There is likely a very thin layer of residual photoresist, called scum, remaining on the silicon surface. Such scum is more apparent for deep and narrow pattern structure where developer is at the bottom of the pattern structure difficult to reach, leaving more scum on silicon surface. The scum, though only a few nanometers, may prevent subsequent pattern transfer process, particularly in wet etching process. Descuming is generally required after photoresist development. The descuming is performed by placing the developed wafer in a plasma etcher for a short-time oxygen plasma etching, normally half minute, to remove the residual photoresist. A caution should be taken that descuming also partially reduces the photoresist thickness and may have detrimental effect on developed photoresist pattern structure.

8. Hard Bake

   Hard bake is to enhance the adhesion of photoresist pattern structure to the silicon wafer by high temperature baking of the developed wafer in oven. Hard bake can also enhance the etch resistance of photoresist patterns. Not all processes require hard bake. It should be aware that hard bake could make photoresist patterns more difficult to remove after pattern transfer.

9. Pattern Transfer

   The purpose of optical lithography is to transfer the mask design onto silicon wafer. Photoresist only serves as a masking layer, i.e., protecting the substrate so that only those areas where photoresist has been developed and substrate surface has been exposed can be further processed, either by etching or by deposition. The pattern transfer techniques will be introduced in Chaps. 7 and 8.

10. Removal of Photoresist

    After pattern transfer, the photoresist has fulfilled its purpose and needs to be removed. Photoresist removal can be a wet process or dry process. Wet process involves various acidic or alkaline or solvent to dissolve the remaining photoresist. The most commonly used solvent is acetone. Acetone can dissolve most of photoresists. Dry process is to remove the photoresist by oxygen plasma etching.

2.3.2 *Characteristics of Photoresists*

Photoresist is the name for all photon-sensitive macromolecule polymers. The name of "resist" indicates its function as an etching-resistant material to protect substrate. They are photon sensitive; therefore, they can be optically defined into pattern structures. Photoresists can be classified into positive tone and negative tone resists, according to the polarity of pattern structures after photon exposure. Figure 2.7 illustrates the difference between positive and negative photoresists. The photochemistry in a positive photoresist is such that long chain of molecules can be broken by photon energy into short chains through a "chain secession" process, while for negative photoresist initially short chains of molecules can join together to form long chains after absorbing photon energy. Macromolecule polymers with long chain are insoluble in developer, whereas polymers with short chain molecules are soluble in developer. Therefore, as illustrated in Fig. 2.7, exposed part is removed by developer for a positive photoresist, while it remains after development for a negative photoresist.

However, there are exceptions that a positive tone resist can create negative tone images. One way is to reverse the tone of mask from dark field (features are transparent on the mask) to a bright field mask (features are opaque on the mask). Another is to use negative tone developer. Standard developer for a positive tone photoresist is alkaline based. Negative tone development (NTD) uses solvent-based developer. As illustrated in Fig. 2.8, the solvent developer reverses the tone of exposed positive resist. The reason to use NTD is that the negative tone photoresists for 193 nm DUV photolithography are always inferior to its positive tone counterparts. Using a positive tone photoresist combined with a bright field mask and NTD can create better quality of exposed positive tone pattern structures, particularly at sub-50 nm resolution [6].

Many types of photoresists have been developed over the years for IC manufacturing. There are four basic constituents in all types of photoresist: resin, solvent, photoactive compound (PAC), and additives. Resin is the main part of a photoresist, which provides the etch-resistant property. Solvent keeps the resin in liquid form so that it can be spin-coated onto a substrate. The PAC determines the photoresist only sensitive to certain wavelength of photon irradiation. For chemically amplified resists (CARs), there is also a constituent called photoacid generator

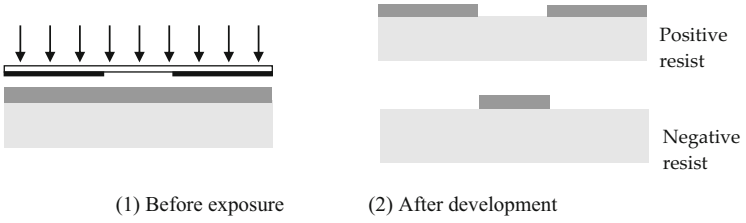


Fig. 2.7 Difference in image polarity between positive and negative tone photoresists

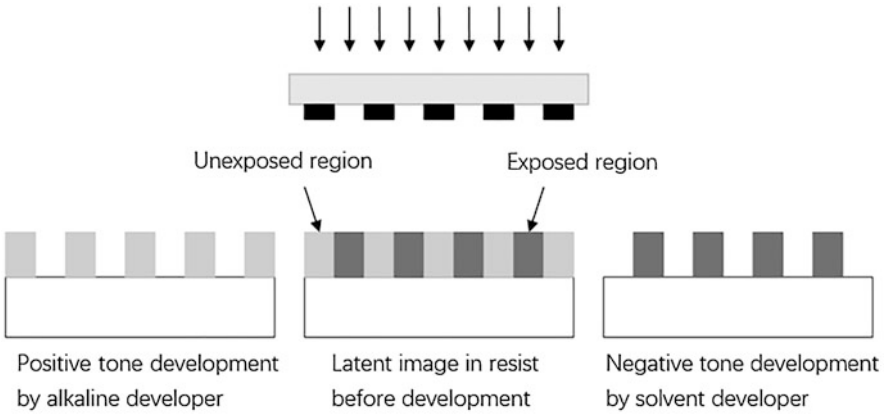


Fig. 2.8 Difference between positive and negative tone development

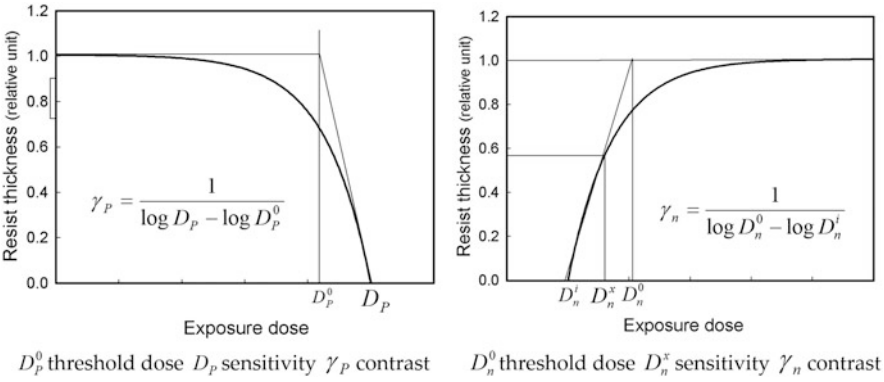


Fig. 2.9 Definition of sensitivity and contrast for positive (left) and negative (right) photoresists

(PAG) which helps to enhance the exposure sensitivity. Additives control light absorbance and solubility of a photoresist. It should be noted that a photoresist may not just be sensitive to photons. Some photoresists are also sensitive to electrons, ions, or X-ray.

A photoresist is characterized by the following parameters:

1. Sensitivity

   Sensitivity determines the exposure speed of a photoresist. The definition of sensitivity is illustrated by the resist development curves which are the remaining resist thickness versus exposure dose, shown in Fig. 2.9. It should be noted that the sensitivity for positive resist and negative resist is defined differently. The higher the sensitivity of a photoresist, the less the exposure dose it needs, hence the faster the exposure. Exposure dose is in the unit of mJ/cm^2.

The chain secession or cross-linking in a photoresist normally takes place upon absorbing photon energy of certain wavelength. However, these processes can take place after exposure and during post-exposure baking for CARs, which enables the CARs to have much higher sensitivity because less photons are needed to expose them [7].

2. Contrast

   Photoresists with high contrast can produce pattern structures with vertical sidewall and high aspect ratio (ratio of pattern height to pattern width). The definition of contrast for both positive and negative photoresists is defined as the slope of development curve shown in Fig. 2.9. The steeper the slope of the curve, the higher the photoresist contrast. Resist contrast has a direct impact on optical lithography resolution.

3. Etch Selectivity

   A high etch selectivity is desired if the photoresist is to be used as a masking layer in pattern transfer by plasma etching. Etch selectivity is defined by the ratio of substrate etch rate and photoresist etch rate. If a photoresist has an etch selectivity of 10 with silicon, it implies for a silicon etch rate of 1 μm/min; the photoresist loss is 100 nm/min. Etch selectivity determines how thick resist layer is needed for a certain etch depth of substrate material.

4. Resolution

   The achievable resolution of a photoresist depends on a number of factors; (a) the resolution of optical exposure system; (b) the molecule weight, molecule weight distribution, and contrast and resist thickness; and (c) development condition, pre- and post-bake temperature, etc. In general, thin resist layer makes it easy to achieve high resolution patterns. However, the etch selectivity has to be taken into account when deciding the resist thickness. Overdevelopment for positive resist and underdevelopment for negative resist will all result in reduced resolution in photoresist patterns. High post-baking temperature will also affect the resolution of resist patterns because of its softening effect on developed resist pattern structures.

5. Exposure Latitude

   If the critical dimension (CD) of exposed pattern changes less when exposure dose deviates from the optimum dose, we say the photoresist has larger exposure latitude. The exposure latitude is best illustrated in Fig. 2.10 where linewidth varies differently with exposure dose for resist A and resist B. The designed linewidth is 0.45 μm. The exposure latitude is quantitatively defined as the exposure dose range within ±10% of CD variation. By comparison, resist A has larger exposure latitude than resist B because resist A allows larger variation of exposure dose while still maintaining CD variation within ±10%. In an ideal situation, the exposure dose is fixed and a substrate receives uniform exposure everywhere. In reality, the light source of an optical lithography system may vary due to external factors. A photoresist with larger exposure latitude will be affected less when exposure energy fluctuates or is not uniform across the wafer.

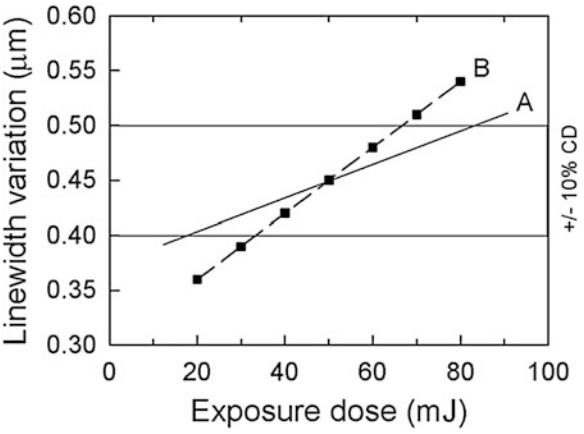


Fig. 2.10 Comparison of linewidth variation with exposure dose variation

6. Process Latitude

   Variations of pre- and post-baking temperatures, development time, concentration of developer, and temperature are all influential over the quality of final resist patterns. Each set of process technology has its optimum conditions. When process conditions deviate from their optima, the performance of a photoresist should vary as little as possible, i.e., having a larger process latitude. A photoresist with large process latitude allows certain fluctuation of process parameters during manufacturing and hence will have higher yield compared to a photoresist with small process latitude.
7. Thermal Stability

   Each photoresist has a glass transition temperature (Tg), above which the resist is in molten state and thermal flow of patterned resist structure may happen. The thermal flow will deform the resist structure and hence deteriorate the pattern quality and resolution.
8. Swelling

   Swelling can happen in some of negative photoresists during development. This is mainly due to the developer molecules that are embedded in the polymer chains, causing volume increase. Swelling can also deform the patterned resist structures.
9. Viscosity

   The viscosity of liquid-form resist determines the flowing ability of the resist. A photoresist with different viscosities will produce different thickness of coating layer, which is important in optical lithography of thick resists. Viscosity can normally be controlled by the solid content of polymer in the resist liquid. The same resist can have different viscosities depending on the polymer concentration. For solvent-based resists, gradual evaporation of solvent after the bottle is opened for some time can increase the polymer concentration and, therefore, increase the viscosity. At the same spin coating condition, fresh resist and old resist may produce different coating thickness due to the change of resist viscosity.

10. Shelf Life

    Every photoresist has its own shelf life, because photoactive compounds may lose its photoactivity with time and solvent may gradually evaporate. Some resists can change property at or above room temperature and must be stored at low temperature. Photoresists are in general sensitive to ambient lights, particularly the UV constituent in the sunlight or artificial light. They have to be contained in a dark brown bottle and optical lithography process line has to be housed in an amber lighting environment.

2.4 Optical Lithography at Shorter Wavelengths

2.4.1 Deep UV

In projection photolithography, the mask is the object to be imaged onto a substrate which is the image plane. The light passing through the mask is the light from the illumination source in an optical lithography system. Therefore, the wavelength appearing in Eq. 2.3 is the wavelength of illumination source. In the early days of optical lithography, mercury lamps were the illumination source. The spectra line at wavelength of 436 nm, called g-line, was first utilized. As the mask features shrank, shorter wavelength at 365 nm (i-line) became the choice of illuminating source. Further demands in imaging small mask features made the mercury lamp unsuitable as the illumination source, because there is simply not enough photon energy at shorter wavelength to meet the requirement of volume production of ICs.

Fortunately, excimer lasers can meet the requirements of both high photon energy and shorter wavelength. Excimer lasers are pulsed gas discharge lasers which produce light output in the ultraviolet region of the spectrum. The commonly used excimer laser sources are listed in Table 2.1, each emitting different wavelength of UV light.

The KrF excimer laser has the highest relative power and was the obvious choice of illumination source to go down the scale of wavelength for volume IC manufacturing. The introduction of excimer laser marked the beginning of deep UV (DUV) optical lithography. The next illumination source for DUV optical lithography was the ArF excimer laser at 193 nm wavelength. It could have gone down even shorter wavelength with excimer lasers. There are the molecular fluorine (F_2) excimer laser emitting photons at 157 nm wavelength and Ar_2 excimer laser at

Table 2.1 Excimer lasers and their relative power

Wavelength	Active gases	Relative power
157 nm	Molecular fluorine (F_2)	10
193 nm	Argon fluoride (ArF)	60
248 nm	Krypton fluoride (KrF)	100
308 nm	Xenon chloride (XeCl)	50
351 nm	Xenon fluoride (XeF)	45

Table 2.2 Reduction of wavelengths in optical lithography

Minimum feature size (nm)	7000–1000	1000–350	350–180	180–32	32 and below
Lithography wavelength (nm)	436 (g-line)	365 (i-line)	248 (DUV)	193 (DUV)	13.5 (EUV)

126 nm wavelength. The semiconductor industry did start concerted effort from 1998 to develop 157 nm optical lithography technology, aiming at IC feature size of 70 nm.

For optical lithography at 157 nm wavelength, there are three critical technology issues: material for making optical lenses, transparent and radiation-durable pellicle for masks, and photoresists. All these technical barriers are stemmed from high absorption of photons at short wavelength such as 157 nm or below. The lens materials at 248 nm and 193 nm wavelengths used to be fused silica, which is no longer suitable at 157 nm because of the high absorption or poor transmission. The high absorption nature at 157 nm wavelength has also prevented the use of pellicle which is a thin film stretched on a frame to protect a photomask from particle contamination. The same problem applies to photoresists. Resists designed for 248 nm and 193 nm lithography are too absorptive at 157 nm. New resist materials have to be developed.

Given the huge efforts needed to make the 157 nm wavelength optical lithography production worthy and the fact that 157 nm can only last one generation of IC manufacturing, the semiconductor industry had to ponder whether to pursue 157 nm or to go directly to much shorter wavelength which is the extreme UV (EUV) at 13.5 nm. The picture became clear in 2004 when immersion optical lithography at 193 nm started to mature and could well take over 157 nm wavelength for 70 nm IC manufacturing or even below [8]. Optical lithography at 157 nm became unnecessary. On the other hand, EUV can play a key role in several generations of IC ahead from 32 nm to 22 nm or below, because of its much shorter wavelength. Table 2.2 has listed the minimum IC feature sizes in the history of semiconductor industry development, and the corresponding wavelengths of optical lithography. There is a big jump from 193 nm to 13.5 nm. Entering 13 nm wavelength regime, optical lithography has come to a completely new era.

2.4.2 Extreme UV

EUV at 13 nm wavelength is in strict sense no longer optical irradiation. It is often called soft X-ray. To differ it from hard X-ray lithography, the semiconductor industry still considers EUV lithography (EUVL) as an optical lithography. In many aspects, EUVL retains the look and feel of optical lithography as practiced so far. However, conventional refractive optics ceases to function at EUV because of its strong absorption in virtually all materials. Instead, reflective optics has to be used

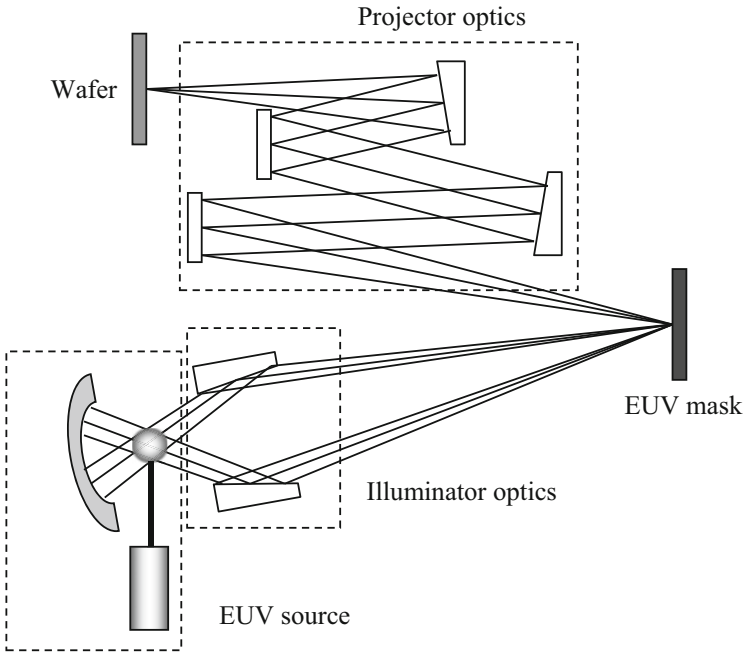


Fig. 2.11 Schematic of EUV lithography system

for both mask and focusing optics. The basic system of EUV lithography is schematically shown in Fig. 2.11. It shows that the EUV radiation from an EUV source is collected by a set of reflective mirrors (illumination optics) and projected onto a reflective mask. The reflected mask image is then focused by another set of reflective mirrors (projection optics) and projected onto a substrate to expose the EUV resist layer. In addition, EUV imaging must be carried out in a near vacuum condition to reduce atmospheric absorption of EUV light.

There are four critical components in EUVL: EUV source, EUV projection optics, EUV mask, and EUV resist. The state-of-the-art developments of the four critical components are described as follows.

2.4.2.1 EUV Source

EUV radiation can be generated by two methods: plasma and synchrotron radiation source. Synchrotron radiation source (SRS) is a large-scale instrument which can only be built and owned by a national government and is mainly for fundamental scientific research. Industrially dedicated EUV sources for lithography purpose are mostly based on plasma-generated EUV radiation. There are two types of plasmas that can generate EUV wavelength radiation: laser-produced plasma (LPP) and discharge-produced plasma (DPP). Some elements such as Li, Xe, and Sn, after

being ionized, can emit short wavelength photons when electrons in the atomic structure jump from higher energy band to lower energy band. For example, ionized Li atom has a radiation spectrum from 10 to 22 nm, with the highest intensity of spectra line at 13.5 nm [9]. This has been the reason that the central wavelength for EUVL is fixed at 13.5 nm. After years of investigations, LPP of liquid Sn droplets became the prevailing choice of EUV source, and maximum 6% conversion efficiency has been achieved [10]. ASML subsidiary Cymer and Japanese company Gigaphoton are currently the major players in the EUV source technology. In the last 10 years, the EUV source power has been steadily increasing, from 40 W in 2014 to 250 W in 2017, achieving 125 WPH (wafer per hour) exposure throughput in the ASML's NXE:3400B EUV scanner [11]. Later in the NXE:3400C scanner, 250 W/170 WPH at 30 mJ/cm^2 exposure dose was achieved [12]. The industry has now made progress in developing 600 W source and in the next steps toward even higher powers [13].

2.4.2.2 EUV Optics

The key component of EUV optics is the multilayer reflective mirror. EUV reflectivity of any single materials at near-normal incidence is very low. In order to achieve reasonable reflectivity near-normal incidence, the mirror surface must be coated with multilayer thin film coatings known as distributed Bragg reflectors. These coatings consist of a large number of alternating layers of materials having dissimilar EUV optical constants, and they provide a resonant reflectivity when the period of the layers is approximately $\lambda/2$ of EUV wavelength. The most developed and best understood reflective multilayers for EUV are made of alternating layers of Mo and Si, and they have the highest reflectivity at wavelengths of about 13 nm. Figure 2.12 shows the theoretical reflectivity of a mirror consisting of 50 pairs of Mo (2.76 nm)-Si (4.14 nm) multilayer at normal incidence of EUV [14]. In practice,

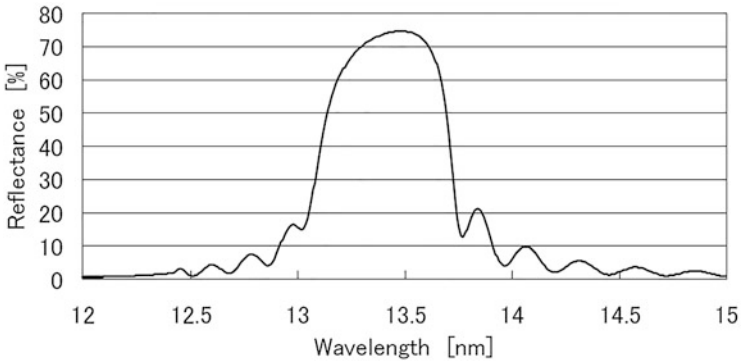


Fig. 2.12 Theoretical reflectivity of a mirror consisting of Mo-Si multilayer at normal incidence of EUV radiation

around 68% reflectivity can be achieved by a multilayer Mo-Si mirror. The causes of decrease in reflectivity are absorption of EUV energy by impurities in the deposited film, index change due to formation of interdiffusion of layers, and scattering due to interface roughness. The major issues for EUV optics are the contamination control and lifetime under EUV irradiation. The latest development in EUV optics is to increase the numerical aperture (NA). Current EUV scanners in use have the NA of 0.33. The next generation of scanner developed by ASML has the NA of 0.55, which is capable of single-exposure at sub-8 nm half-pitch resolution for 2 nm logic node of ICs [15]. In December 2023, ASML shipped its first high-NA (0.55) EUV scanner to Intel, and next generation of hyper-NA (>0.7) EUV scanner is being developed according to the ASML's 2023 Annual Report.

2.4.2.3 EUV Mask

EUV mask works the same way as the EUV mirror. It has to be reflective. There are two stages of processes involved in making EUV masks: first is the making of mask blank and second is the patterning of absorber layer. Figure 2.13 shows the cross section of an EUV mask. The blank is similar to an EUV mirror except that there is an absorber layer on top of the multiple pairs of Mo-Si layers. The substrate must be a low thermal expansion material with flatness better than 50 nm and free of defects. A number of metals can be used as the absorber layer, such as Al, Cr, Ta, and W, with the Ta-based material to be the favorite [16]. Because EUV projection optics is reduction optics with reduction ratio of 4:1, the patterning of EUV absorber is less demanding. Conventional electron beam lithography and reactive ion etching are adequate for the absorber patterning. The most difficult part in EUV mask manufacturing is the defect inspection and repair [17]. There are two groups of defects, the defects in a mask blank and defects in patterned absorber. As EUV imaging relies on reflective surface of mask, any bumps can cause phase error in the final image. It has been characterized that a 65 nm × 1.5 nm (width × height) bump in the multilayer surface can cause 20% deviation of dimension to a 25 nm line [18]. Defects as small as 50 nm on the absorber patterns can be printable [19]. For repair of defects in conventional optical mask, focused ion beam has been the universally used repair tool. However, focused ion beam repair of opaque defect in

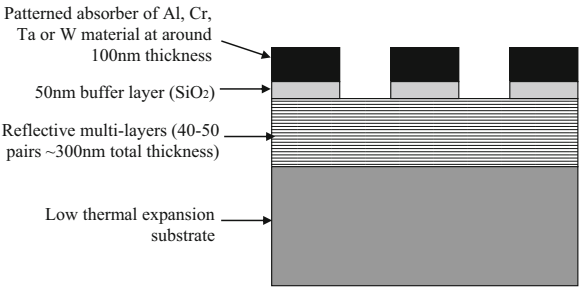

Patterned absorber of Al, Cr, Ta or W material at around 100nm thickness
50nm buffer layer (SiO_2)
Reflective multi-layers (40-50 pairs ~300nm total thickness)
Low thermal expansion substrate

Fig. 2.13 Cross-sectional view of EUV mask

EUV masks can easily cause both damage and gallium staining to the substrate, and gallium is very absorbent at EUV wavelength of 13.5 nm. In this case, electron beam repair method is preferred. Electron beam irradiation assisted by reactive gases can either etch or deposit materials with high resolution and high accuracy, which makes it a perfect tool for defect repair in EUV masks [20].

2.4.2.4 EUV Resists

Resists for EUV lithography are required to have higher sensitivity and low line edge roughness (LER). The traditional chemically amplified resists (CARs) used in DUV lithography are no longer suitable for EUV because they are more transparent at 13 nm wavelength than at 193 nm wavelength. Less absorption of EUV radiation results in low sensitivity. The CARs also have worse LER for features below 22 nm because of the photoacid diffusion process after exposure. For this reason, new resist formulations consisting of atoms with enhanced absorptivity in EUV regime, such as metallic atoms of Hf, Zr, Zn, and Sn, have been proposed as alternative to traditional CARs [21]. In addition to better absorption of EUV photons, the metal-containing inorganic photoresists, which are non-chemically amplified, have another advantage over organic resists. For EUV lithography of sub-32 nm generation ICs, the resist thickness is less than 35 nm. Such thin layer of resist has to have good etch selectivity to substrate materials in order to transfer the resist patterns to substrate. The metal-containing inorganic resists have much better etch resistance in most plasma etching processes [22]. An all-dry processed negative tone EUV resist with resolution down to sub-10 nm lines was also demonstrated, which can be coated and developed by vacuum sublimation [23].

After years of efforts, EUV lithography reached a significant milestone: readiness for high-volume production of advanced semiconductor devices. An EUV source power of 250 W has been realized, providing a tool throughput capability exceeding 140 wafers per hour at a dose of 20 mJ/cm^2. The full-wafer critical dimension (CD) uniformity is less than 0.5 nm for multiple systems and matched-machine overlay is at 1.1 nm [24]. 2019 was the year the EUV patterning technology was deployed for the first time for the mass production of logic chips of the 7 nm technology generation.

2.4.3 X-Ray

The ultimate short wavelength photon radiation would be the X-ray. X-ray is the radiation wavelength spanning from 1 nm of the soft X-ray to 0.1 nm of hard X-ray. The fundamental difference between X-ray wave and optical wave is that X-ray can penetrate the majority of materials. Only materials with high atomic number can absorb X-ray. X-ray cannot be focused because the refractive index of all materials to X-ray is the same ($n \approx 1$). Therefore, even if X-ray enters from free space to a

material or from one material to another, there is no refraction. X-ray exposure of mask image can only be 1:1 with a proximity gap between mask and resist surface. The opaque and transparent features on an X-ray mask are defined by heavy metal absorber and thin silicon membrane (1–2 μm thickness).

Because of its short wavelength, X-ray lithography once was a very hopeful candidate technology to replace optical lithography for IC manufacturing. IBM started the technology development in 1980. Motorola, NEC, NTT, and Mitsubishi all had invested in the technology. The beginning of and mid-1990s were the heydays of X-ray lithography R&D. As early as in the mid-1990s when optical lithography was still struggling to get down to 0.25 μm circuit dimension, X-ray lithography had already achieved 0.12 μm experimental DRAM chips and 0.1 μm logic chips [25]. However, X-ray lithography eventually did not succeed to replace optical lithography. The first hurdle is the fabrication of mask. As X-ray exposure is 1:1 pattern transfer, the heavy metal absorber has to have the same dimension as the resist feature. In addition, local stress caused by heavy metal absorber and internal stress of supporting membrane all can have influences on the pattern resolution.

The second hurdle is to accurately control the mask-wafer gap. High resolution X-ray lithography requires the gap to be controlled within 10 μm. Variation of the gap will cause variation of image resolution. This requires the flatness of silicon wafer to be within ±0.25 μm and flatness of the mask within ±0.5 μm [26]. These are very stringent requirements for a large-size silicon wafer and a large area of membrane. The gap control is also required to be within ±0.5 μm accuracy. Bearing in mind that the mask membrane is only 1–2 μm thick and very easy to crack, it makes the gap control a lot harder.

The third hurdle is the photoelectrons and Auger electrons generated by high energy X-ray irradiation in resist. These low energy electrons are more efficient to interact with resist polymer molecules, causing exposure of the resist. Upon acquiring energy from X-ray, these electrons scatter to some distance within the resist layer. The consequence of electron scattering is the expansion of exposure pattern dimension, hence degradation of lithography resolution.

On top of all the technical challenges is the economical factor. For industrial-scale manufacturing, the X-ray "light" has to be from a synchrotron radiation source (SRS). Construction of a large-scale SRS costs over $100 million or a compact size SRS costs about $20 M. Although one SRS can have 10–20 beam lines equivalent to 10–20 optical steppers, they all reply on proper working condition of the SRS. If the SRS breaks down, all the beam lines will stop working. Had the optical lithography not kept the pace of shrinking down in IC feature dimension, the semiconductor industry would have had no choice but to adopt X-ray lithography. However, optical lithography has been continuously advancing. New resolution record has been broken time and again beyond anyone's dream decades ago. Fabrication of optical masks is much simpler and easier than fabrication of X-ray masks. X-ray lithography has eventually failed to win favors from the semiconductor industry to become the mainstream manufacturing technology [27]. X-ray lithography is now mainly used not for high resolution but for high aspect ratio features because of its powerful penetration capability [28], particularly for the LIGA process [29].

2.5 Optical Lithography at High NA

The second parameter influencing the resolution of optical lithography, in addition to the wavelength as seen in Eq. 2.3, is the numerical aperture (NA). As explained earlier, NA is equivalent to the diameter of a lens. While lights passing through a small mask feature (transparent opening) diffract in wide angle as illustrated in Fig. 2.2, a lens with large diameter can collect more diffracted lights and focus onto the image plane. It is crucial to collect the wide-angle diffracted lights because they contain the information about the details of the mask feature. The light diffraction through a small feature has the characteristics that the finer the details of the feature, the further the lights diffract from the central optics axis. Or in terms of Fourier optics, smaller spatial detail can be referred to as being of higher "spatial frequency." Such characteristics can be best explained in Fig. 2.14a, b.

Figure 2.14a is square wave function, representing slits or a series of line features in a mask, which have the ideal distribution of light intensity (without diffraction). After diffraction at the image plane, the square wave function is transformed into a series of sinusoidal wave functions ranging from the base spatial frequency ($m = 1$ in Eq. 2.1) to higher orders ($m > 1$ in Eq. 2.1). The superimposed sine waves of different spatial frequency become the image of original square function (object) at the image plane. The more the high spatial frequency waves are included, the better the final image approximates the original object, as shown in Fig. 2.14b. As seen in the plots, the superimposed image of up to 200 sinusoidal waves is almost identical to the original square wave function. It is therefore understandable that only an optical lens system with larger NA can allow diffracted lights of much higher spatial frequency to reach the image plane, resulting in better image of the mask features. High NA also means more lights coming through the lens system, hence a brighter image on the image plane. Computer simulation of optical imaging confirms the influence of NA, as shown in Fig. 2.15, where a group of 0.2 μm lines is imaged

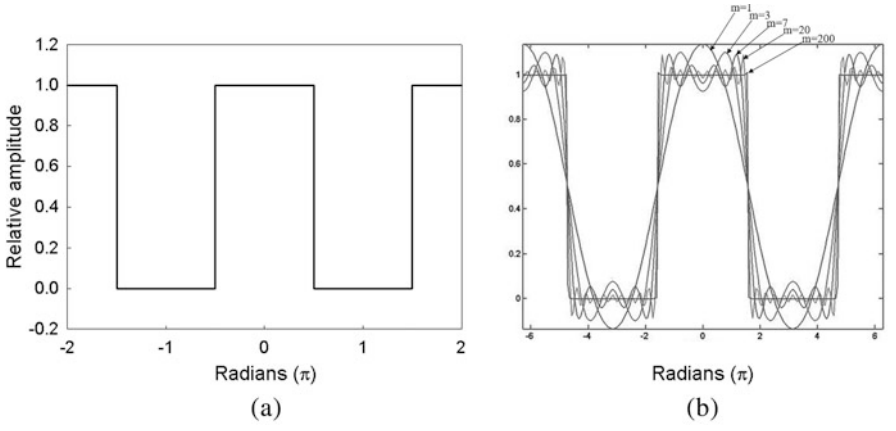


Fig. 2.14 Mask images (**a**) without diffraction and (**b**) with diffraction

through a projection optical system at 248 nm wavelength with three different numerical apertures (NA = 0.5, 0.6, 0.8). The lines imaged at high NA are not only resolved better (high image contrast) but also brighter (high image intensity).

Although it is so obvious that high NA can improve the resolution of optical lithography, resolution is only one requirement for high-quality optical lithography. Another important requirement is the depth of focus (DOF), or commonly known as depth of field in photography. DOF is also related to the wavelength of illumination and numerical aperture of an optical system by Eq. 2.4, where k_2 is a constant ($k_2 < 1$) related to a specific lithography system and photoresist process.

$$\mathrm{DOF} = k_2 \frac{\lambda}{(\mathrm{NA})^2} \tag{2.4}$$

Equation 2.4 indicates that increasing NA has an effect on reducing the DOF. Figure 2.16 compares the image intensity of the same line features as in Fig. 2.15 at NA = 0.8 but with defocus. If there is a defocus, the high NA results in much worse degradation of image quality, as seen in Fig. 2.16.

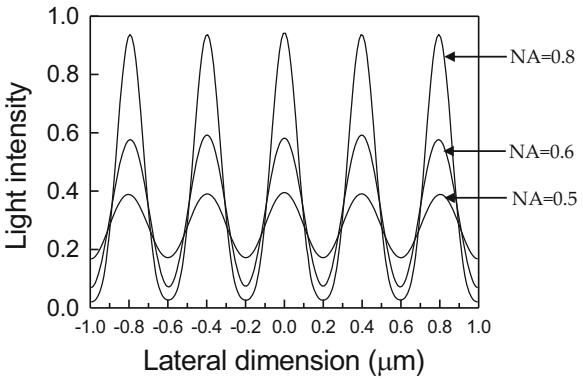


Fig. 2.15 Influence of numerical aperture on optical imaging

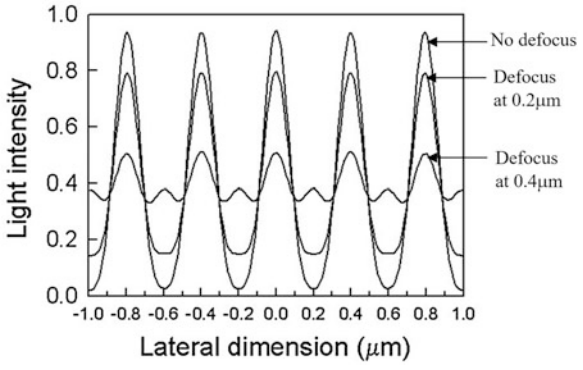


Fig. 2.16 Influence of depth of focus on optical imaging

The DOF defined by Eq. 2.4 is only valid for NA less than 0.5. For higher NA, the equation has to be modified and take the form of Eq. 2.5 (where k_3 is a constant similar to k_2), which causes DOF to drop much more steeply at high NA value [30]

$$\mathrm{DOF} = k_3 \frac{\lambda}{2\left(1 - \sqrt{1 - \mathrm{NA}^2}\right)} \tag{2.5}$$

It must be pointed out that DOF is actually a more important parameter in the lithography of ICs. ICs are always fabricated on large size of silicon wafers, normally in 8″ or 12″ diameter. The wafers in such size are impossible to be perfectly flat, plus there is topography formed by circuit structures in previous process steps. One can imagine that if the DOF is very small, the optical imaging can only be in focus in a very small range of height variation. There would be many places in a wafer where defocus would be inevitable because of the topography or wafer bowing.

Every layer of 2D circuit layout on an IC chip has a group of critical minimum features which dimension is called "critical dimension," or CD. This CD has to be controlled within a certain range (less than ±10% variation from nominal dimension) during fabrication to ensure the final circuit working properly. For example, for a 100 nm critical feature dimension, any errors during fabrication should increase or reduce the feature dimension no more than 10 nm. The DOF of a lithography system has to be large enough to ensure the CD variation is within ±10% everywhere on a wafer. For scientific experiment, resolution may be the only important target to achieve, as the researchers are only interested in a small device area where there are a few pattern structures with minimum dimensions. For large-scale production, high yield can only be possible when all the minimum circuit features across the entire wafer are controlled within the target CD variation. In this case, the highest resolution is only meaningful within certain DOF.

The increase in NA used to be restricted to ensure the DOF is above 1 μm at a specific wavelength, which would give much ease to IC processing. With advances in all aspects of optical lithography, particularly in wafer flatness, introduction of autofocus and auto leveling systems, thinner resist processes, and reduced surface topography via chemical mechanical polishing (CMP), the requirement on DOF has been gradually reduced to below 0.5 μm. This has enabled gradual increase of NA over the years. Figure 2.17 shows the evolution of NA in optical lithography from the 1980s up to mid-2000s. While the wavelength has increased from g-line to deep UV at 193 nm, the NA also increased from 0.3 to above 0.8. By 2009, the DUV TWINSCAN system made by ASML reached 0.93NA which was the highest NA that could be possible [31]. The higher the NA, the more complex, bigger, heavier, and more expensive the lens has become. In the 1980s, a g-line stepper lens at 0.35NA weighed only 14 kg. By the mid-1990s, an i-line stepper lens at 0.63NA weighed over 500 kg [32].

Increasing NA above 0.8 not only encountered the difficulty in lens design/building and left optical lithography with very small DOF but also had the problem

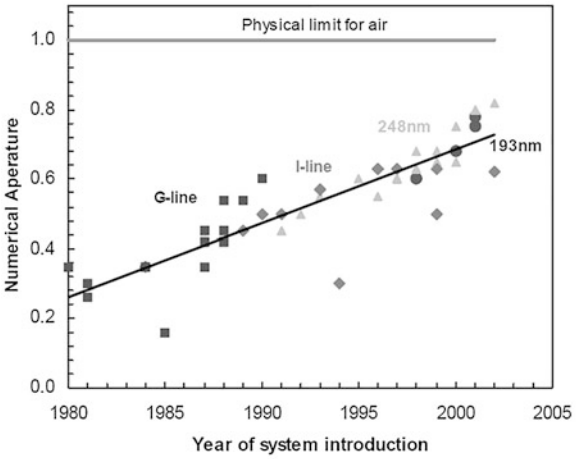


Fig. 2.17 Evolution of numerical aperture in projection optical lithography. (Reprint from Ref. [33] with permission)

of polarization effect at high oblique incidence of light, which degrades the image contrast [34]. High NA has forced optical lithography into an optical imaging regime where the approach angles of the imaging rays are so wide that the light coming out of the lens can no longer be considered as a scalar. Instead, one must consider the vector components of light and take into account the separation between the electric and magnetic field components of light, the so-called s-polarization and p-polarization. When there is incident light in a large oblique angle from air into a polymer resist layer, part of the light (s-polarization) can be reflected from the surface if the angle is greater than Brewster's angle. Reflection of polarized light will reduce the photon energy that goes into the photoresist, causing low contrast of image and low image intensity. Brewster's angle θ_B at which the polarization reflection occurs is expressed as Eq. 2.6, when light from one medium with refractive index n_1 goes into another medium with refractive index n_2. In the case of optical lithography, light goes from air ($n_1 = 1$) to a photoresist layer (n_2).

$$\tan(\theta_B) = \frac{n_2}{n_1} \tag{2.6}$$

High refractive index photoresist will be able to tolerate imaging with high-NA optics. For example, if a photoresist has refractive index of 1.5, Brewster's angle θ_B will be 56°, which translates into NA of 0.83 based on Eq. 2.2. Theoretically, further increase in NA can be tolerated by increasing the refractive index of photoresist, ignoring the difficulty and cost of making high-NA lenses. However, the refractive index of photoresist is not a property that can be easily adjustable. One way to solve the polarization problem is to illuminate the mask with a deliberately polarized light source, for example, illumination containing only the p-polarization light. In this way, there will be no reflection of light from the photoresist surface, even with very high-NA optics. Improvement in DOF and lithography quality has been observed using polarized illumination at lens NA of 0.92 [35].

2.6 Immersion Lithography

The most important development in the history of optical lithography is the introduction of immersion imaging, the so-called immersion optical lithography [36]. Immersion imaging scheme came just at the time when optical lithography has hit the limit at 193 nm wavelength and the semiconductor industry is facing the choice of whether to go for the problematic 157 nm option or to go for the so-called next generation lithography (NGL) which includes EUV, electron projection, ion projection, and nanoimprinting.

It is obvious that if the refractive index n in Eq. 2.2 is not for air ($n = 1$) but for a liquid ($n > 1$), then the NA can be more than 1. The immersion imaging, also called "wet" imaging, is not a new concept. The medical and biological research has for a century used immersion imaging in optical microscope to enhance the image resolution. The wet imaging is illustrated and compared with conventional "dry" imaging in Fig. 2.18. The liquid between the objective lens and imaging plane has effectively reduced the angle of light coming out of the lens. Although the NA of optical lens is the same as for dry imaging, the effective NA seen from the imaging plane side has increased to a value greater than actual NA of the lens. Table 2.3 shows the angle of refracted light in different mediums for a lens with a fixed NA. The light, which used to be entering photoresist at 58° angle in the case of dry imaging, is now entering at 36° angle, equivalent to the NA value of about 1.22. However, the incident light angle cannot be increased by merely placing an immersion fluid between an objective lens and the image plane. The objective lens has to be specifically designed to allow larger angle of light interfacing with the immersion

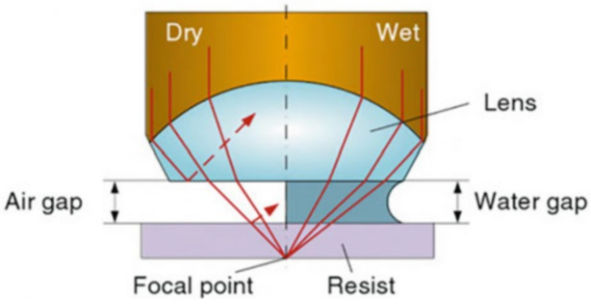


Fig. 2.18 Schematic comparison of "wet" (right half) and "dry" (left half) imaging

Table 2.3 Refractive angles of light in different mediums at NA = 0.85

Medium	Refractive index (n)	Angle of light (θ)
Lens	1.55	33°
Air	1	58°
Water	1.437	36°
Photoresist	1.7	30°

fluid. The most frequently used form of an immersion lens is a prism or hemispherical element, as the one shown in Fig. 2.18.

Immersion imaging at 193 nm (193i) has indeed made a leap forward in terms of sub-100 nm IC manufacturing using optical lithography. Its advantage to extend further into sub-50 nm resolution regime is obvious. For example, assuming $k_1 = 0.25$ and the numerical aperture of lens is 0.75, the dry lithography at 193 nm wavelength can achieve resolution of 65 nm, whereas the same optical system with immersion imaging using water can achieve resolution of 45 nm. While immersion imaging is such a simple and mature concept, why did it take so long for the semiconductor industry to seriously consider the merits of replacing air with a liquid? This was because to realize immersion optical lithography, there were many technical difficulties. In the past, it was not worth the effort to solve these technical difficulties, as conventional optics was more than adequate to meet the resolution requirement of IC production above 100 nm. As optical lithography was approaching its end of life, any measures that could revive it would be seriously considered.

One of the most serious technical difficulties is how to move the wafer stage rapidly relative to the objective lens while in water without stirring up bubbles. In order to achieve high throughput, the wafer stage must accelerate rapidly and move accurately to each exposure field location, then settle, and scan the image before stepping to the next location, all in a short period of time. Maintaining a consistent bubble-free liquid between the lens and the wafer is very difficult. There are basically three approaches to solving the problem [33]. The first approach is to submerge the whole chuck, wafer, and lens in a pool of water. The issue with the pool approach is that a complex system of servomotors and laser interferometers is required to accurately move the chuck. Submerging this whole system would require significantly complex engineering. The second approach is to limit the pool size to the top of the chuck. This technique would keep all of the chuck control mechanisms dry but would add considerable mass to the chuck that must rapidly accelerate. The third and currently being implemented technique is to dispense the water between the lens and the wafer with a nozzle and rely on surface tension to maintain a water "puddle" while the wafer stage moves to each die location during step-scan process. Because the water puddle moves with the lens, no bubbles are stirred up during the step-scan process.

Another critical issue is the defects that are likely to form on photoresist surface due to water marks and drying stains [37]. Conventional 193 nm photoresists that are optimized for "dry" exposures may not be suitable for immersion exposure since resist components show some solubility in water and thus leach out, thereby contaminating the water and creating defects. The contaminated water can further degrade the front lens and wafer stage of the scanner. Water can also permeate into the resist film, causing resist swelling and changing its photochemical properties. An additional topcoat layer is an effective solution to block both the leaching of resist components into the water and the water permeation into the photoresist. The topcoat is spin-coated on the resist and is transparent at 193 nm wavelength. The topcoat layer is removed after exposure and post-exposure bake (PEB) but before pattern

development. There are two types of topcoats: those only soluble in some solvent and those that dissolve in resist developer (aqueous trimethyl ammonium hydroxide (TMAH)). The trend is toward developer-soluble topcoats, because the additional process complexity associated with a solvent-removed topcoat is too high. As 193 nm immersion lithography has been introduced to volume production of ICs, many technical issues associated with topcoat are being investigated. Whether a topcoat is needed is still debatable [38]. In the meantime, non-topcoat photoresists have been developed, since the additional topcoat will inevitably complicate the resist process and increase the cost of manufacturing [39, 40].

The immersion imaging using water ($n = 1.44$) can achieve NA at 1.35 and is capable of 45 nm resolution with 193 nm wavelength. This is about the limit with water as the immersion liquid. Further improvement in resolution has to seek an immersion liquid with higher refractive index ($n = 1.8 \sim 2.0$). At present water has similar refractive index as the glass as shown in Table 2.3. To maintain index matching, high refractive index glass material is also needed for making the objective lens. In 2006 IBM demonstrated 29.9 nm line and space features achieved by 193i lithography using a high index fluid ($n = 1.64$) and a quartz crystal prism ($n = 1.67$) at the end of objective lens to match the high index fluid [41].

The optical glass for making lenses and water as the immersion fluid are all well known for their physical properties. Introduction of new materials for both the immersion liquids and lens materials will lead optical lithography into unknown territory. For high index immersion lithography to enable 1.70NA imaging, a high index lens element with an absorbance <0.005/cm, a fluid with an index of ≥ 1.80, and a resist with an index >1.9 are needed [42]. Though high index immersion lithography was attractive, the IC industry did not pursue down this road, not because there were fundamental obstacles which were hard to surpass but because the industry wanted something which could be quickly adopted for 32 nm and below. The alternative technology is the multi-patterning method which will be described in Sect. 2.7.5 of this chapter. Since 2009, the 193i lithography has been staying with water as the high index liquid [42].

2.7 Optical Lithography at Low k_1 Factor

The third parameter influencing the resolution of optical lithography, as seen in Eq. 2.3, is the k_1 factor. The k_1 factor encompasses any measures that can improve the resolution of imaging, not only in the optics but also in masks and photoresist layer. Apart from developing shorter wavelength illumination and high-NA optics as introduced in previous sections, any other efforts in the development of optical lithography over the last 40 years have been concentrated on lowering the k_1 factor. Figure 2.19 shows the reduction of k_1 factor from g-line optical lithography to EUV lithography [43], which pushed the IC feature dimension from 1 μm in the 1980s to below 5 nm node at present day.

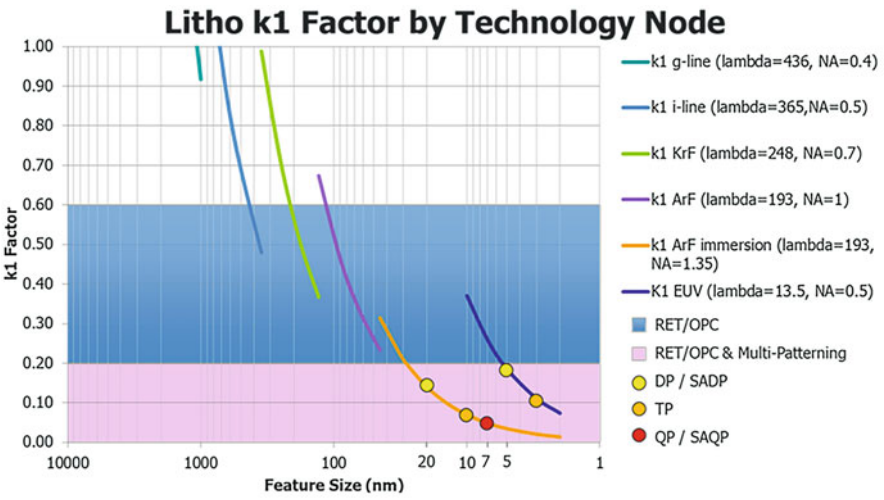


Fig. 2.19 Reduction of IC feature size enabled by reduction of k_1 factor. (Courtesy of David Abercrombie, Semiconductor Engineering [43])

Since the optical lithography is based on diffraction optics, there is a theoretical resolution limit. Traditionally, the resolution limit is defined by Rayleigh criterion. A recent article argued that Abbe diffraction limit gave a more accurate description [44]. In 1873, the German physicist Ernst Abbe discovered that the resolution of a microscope is not controlled by how good the microscope can be but by the wavelength of light used and the aperture of its optics. A microscope cannot resolve two objects located closer than λ/2NA, which translates into Eq. 2.7 where R is the distance between the centers of two separate line opening, λ is the wavelength of light, and NA is the numerical aperture of the imaging lens.

$$R = 0.5\frac{\lambda}{\mathrm{NA}} \tag{2.7}$$

Before 2000, the technology nodes in IC industry were defined by the minimum achievable circuit feature dimension, i.e., linewidth. For example, when speaking of 0.35 μm, 0.25 μm, or 0.18 μm IC technologies, they implicate the minimum linewidth of circuit feature in an IC chip. However, minimum linewidth cannot represent the true integration level of an IC because it does not indicate how dense the circuit features are in an IC. In fact, it is much easier to achieve high resolution for single isolated line than for dense line/space pattern. The true representation of circuit integration density is the pitch, not the linewidth. The pitch is the separation distance between two lines, as shown in Fig. 2.20 which shows two lines of same width but different pitches (P). The simulated optical intensities of two separate lines on a mask shown in Fig. 2.21 demonstrate that different separation distances (pitch) result in different optical images. The smaller the pitch, the worse the resolvability of

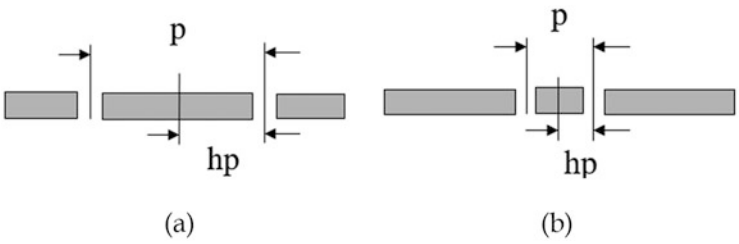


Fig. 2.20 Same linewidth of (**a**) large half pitch and (**b**) small half pitch

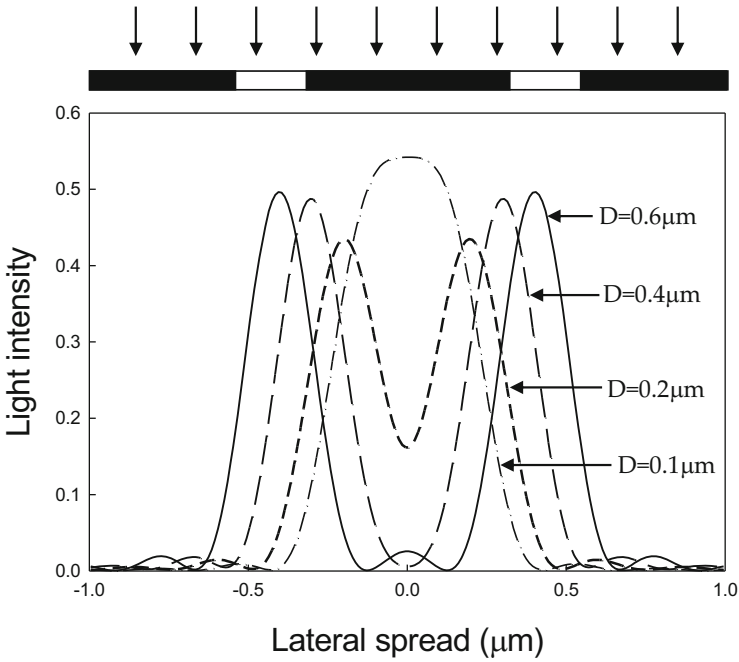


Fig. 2.21 Optical intensity distributions of two lines (0.2 μm linewidth) on mask with different separation distance D (pitch)

the lines even though their linewidth is the same. For this reason, the semiconductor industry started to use "half pitch" (hp) as illustrated in Fig. 2.20, to define the technology node since the 2001 version of International Technology Roadmap for Semiconductors (ITRS) [8]. By this definition, the theoretical limit of k_1 factor would be 0.25 which is shown by the dashed line in Fig. 2.19.

In the early days of semiconductor manufacturing, k_1 factor was much larger. With the continuous shrink of circuit dimension, the resolution of optical lithography was restrained by available illumination wavelength and numerical aperture of

Table 2.4 Summary of resolution enhancement techniques

Components of optical lithography	Innovations to reduce k_1 factor
Illumination optics	Off-axis illumination (OAI): annular, quadrupole, or self-defined
Photomask	Phase-shifting masks (PSMs), optical proximity correction (OPC)
Photoresist	Top surface imaging (TSI), antireflective coating (ARC), inorganic resists
Multiple patterning	Multiple lithography and etch, self-aligned multiple patterning

optical systems. Lowering k_1 factor became the focus of development. There had been innovations in every step of imaging process from the top at the illumination source to the middle at the mask level, and down to the bottom at the photoresist layer. They were all under the name of "resolution enhancement technique" (RET) and contributed to lowering the k_1 factor from 0.6 downward and even breaking the resolution limit ($k_1 < 0.25$), as seen in Fig. 2.19. Table 2.4 summarizes the RETs that have been developed to improve image resolution, contrast, and depth of focus. The following sections will introduce each of these innovative techniques.

2.7.1 Off-Axis Illumination (OAI)

Illumination light through a photomask will be diffracted, as described in Sect. 2.2. The smaller the mask feature, the larger the angle of diffracted light (Fig. 2.2). An optical system has limited resolution because the large angular diffractions are cut off by finite NA of a lens. However, if the illumination is deliberately tilted through an off-axis aperture instead of an on-axis aperture, light passes through the mask feature in an oblique angle, while the zero-order illumination is blocked. Figure 2.22 compares the conventional illumination (a) and off-axis illumination (b). In this way the high orders of diffraction can pass through the mask feature to be imaged. It should then enhance the image resolution.

OAI was first proposed in the early 1990s with two schemes: annular OAI [45] and quadrupole OAI, also called SHRINC (Super High Resolution by Illumi-Nation Control) [46]. Both schemes are illustrated in Fig. 2.22b. Quadrupole OAI is most effective for line and space (L/S) pattern but less for isolated features. With quadrupole OAI, 0.275 μm L/S pattern can be resolved with i-line lithography. Quadrupole OAI is also dependent on the orientation of L/S patterns. It gives the best results for vertical or horizontal L/S pattern but less optimum for L/S with 45° angle. Annular OAI, on the other hand, does not have the discrimination of pattern orientation, but the resolution enhancement effect with annular OAI is rather weak compared to quadrupole OAI.

OAI is the easiest and requires the minimum cost to implement in an optical lithography system among other resolution enhancement techniques (RETs). It

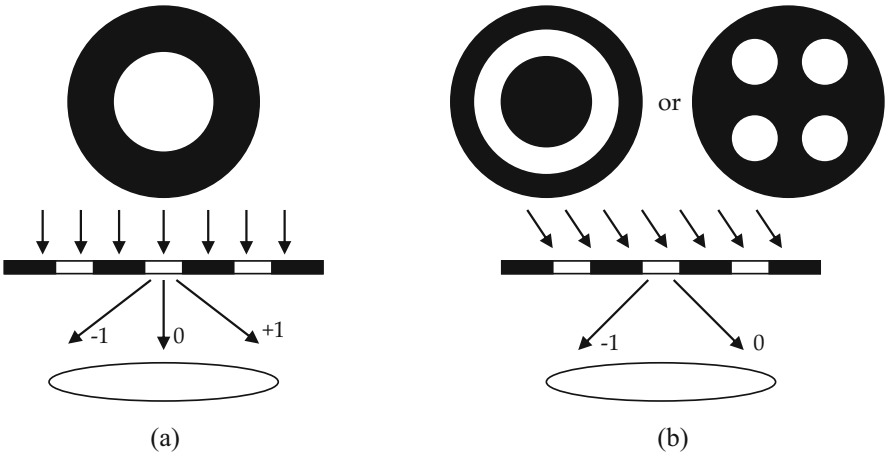


Fig. 2.22 Comparison of conventional and off-axis illumination schemes; (**a**) conventional aperture, (**b**) off-axis apertures (annular or quadrupole)

certainly is the most widespread used RET in commercial optical lithography systems. Because the OAI effect is dependent on pattern density and orientation, the state-of-the-art OAI apparatus are programmable [47]. It can be tuned to a particular set of mask patterns to achieve the optimum performance. For example, the state-of-the-art DUV scanners made by ASML are equipped with a FlexRay illuminator which has 4000 micromirrors, and the illumination conditions can be self-defined [48]. This is particularly useful as the mask patterns become increasingly complex due to the incorporation of many phase shift and proximity correction features. The illumination can be adapted to particular mask patterns through the source-mask optimization (SMO) [49], which has become a standard practice in IC design process.

2.7.2 Phase-Shifting Mask (PSM)

The concept of phase-shifting was proposed in 1982, whereby the light diffraction of a mask feature can be suppressed by introducing light wave with 180° phase difference surrounding the mask feature [50]. Figure 2.23 schematically shows how the light intensity distribution from a mask feature can be made narrower by introducing light of opposite phase in comparison with imaging of a conventional binary photomask. There are a number of ways to realize phase-shifting on a mask, depending on where and how to place the phase shifters. Figure 2.24(i) shows five different types of PSM in comparison with a conventional photomask. The shaded region on a PSM is 100% transparent to light but with 180° phase difference from unshaded pattern region, except the attenuated PSM (Fig. 2.24i–f) where the shaded

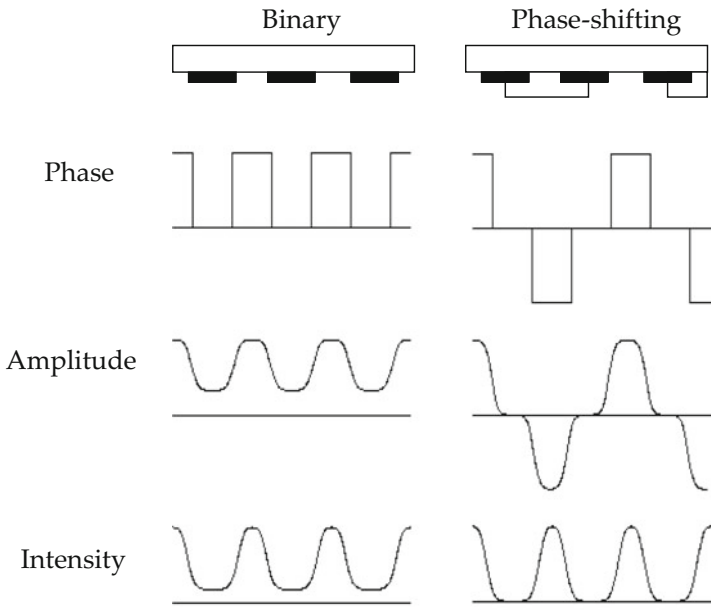


Fig. 2.23 The concept of phase-shifting mask

region is only 5–15% transparent to light but has 180° phase-shifting. Figure 2.24 (ii) conceptually compares the phase and intensity produced by binary and PSMs. Different phase-shifting methods have different performance on resolution enhancement, with the alternating PSM and chromeless PSM being the strongest phase-shifting methods which produce the greatest improvement on imaging resolution. Figure 2.25 shows the simulated light intensity distributions (aerial image) of L/S pattern for conventional mask, alternating PSM, and attenuated PSM. The linewidth is 0.2 μm and exposed with 248 nm wavelength and 0.5NA. It can be seen that the alternating PSM has the most significant improvement in aerial image over the imaging of conventional mask. The attenuated PSM has improvement over conventional mask but not as pronounced as the alternating PSM.

Although the alternating PSM has superb performance in resolution enhancement, there are three critical issues with this type of PSM. First, alternating PSM is only suitable for patterns with high periodicity. They are very difficult to apply to a random design layout. Some memory cell layouts have been modified to be ideally suited for this approach, but this is not possible for logic applications like microprocessors or application-specific integrated circuits (ASICs). There are dedicated software tools which can deal with placing phase-shifting patterns to random design layout, but not all issues have been solved.

The second issue is the phase conflict. Each 180° phase-shifting region has a boundary. If the phase shifter is adjacent to a transparent 0° phase region, the transition between regions of 180° and 0° phase will always print out as a dark

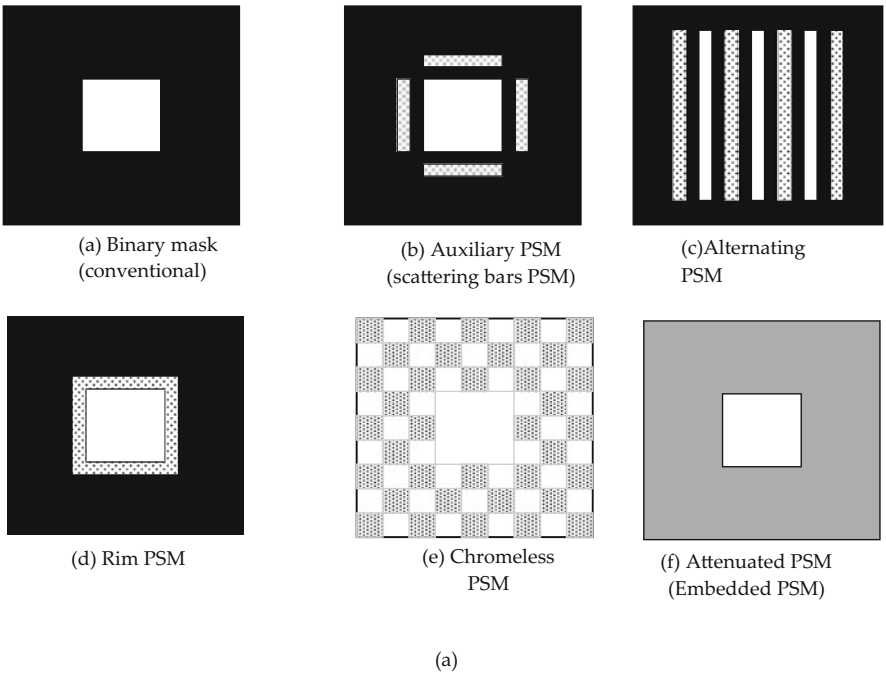


Fig. 2.24 (i) Layouts of different PSM schemes and (ii) phase, amplitude, and light intensity comparison. (Reprint from Ref. [51] with permission)

narrow line whether this is intended or not, because of the cancellation of light at the transition region. For most of IC designs, it is very difficult to avoid these undesirable features created by the phase conflict. There are two ways to overcome this problem. One way is to apply a second exposure with a trim mask to eliminate these dark lines. This approach will increase the exposure time and require an additional mask. Another way is to design some intermediate phase shifters at the transition

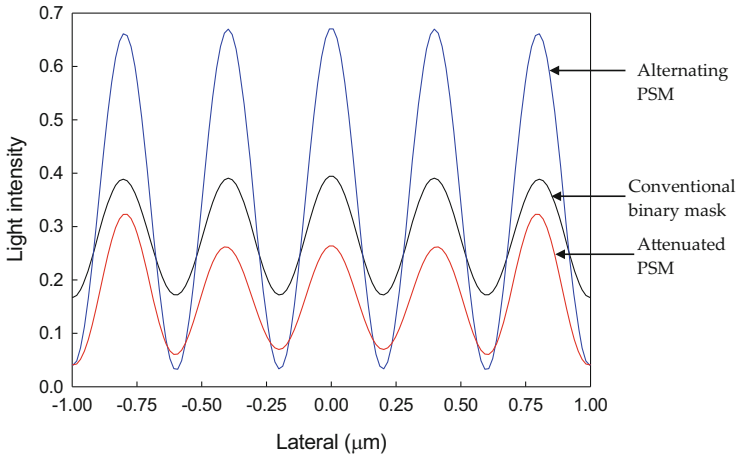


Fig. 2.25 Comparison of aerial images of conventional mask with alternating and attenuated PSMs

region so that the transition of phase is not from 180° abruptly to 0°, but to 120°, 60°, and 0°, for example. Both methods require extra space around the critical features and hence are less suitable for densely packed pattern layouts, such as memory cells.

The third issue is mask manufacturing complexity. Phase-shifting layer has to be produced by additional exposure and deposition or etching of the mask. This additional manufacturing step not only increases the mask cost but also requires accurate alignment of two different patterns and increases the possibility of defects. Inspection and repair of phase defects are complicated. The overall cost of a PSM is at least ten times higher than a conventional binary mask.

Although alternating PSM has the best performance, the first and most widely used PSM in optical lithography for IC manufacturing was the attenuated PSM, solely because it is relatively easy to make and of low cost. Attenuated PSM (attPSM) is also called half-tone PSM or embedded PSM. As shown in Fig. 2.24i–f, the attenuated PSM is not much different from a conventional binary mask (Fig. 2.24i–a), except that the chrome layer on a conventional mask is replaced by a thin film which gives 5–15% light transmission with 180° phase shift. The transmission of attenuation film has to be carefully controlled because of the side lobes which become unwanted features in photoresist [52]. Attenuated PSM is particularly suitable for isolated or low density features, which makes it the best phase-shifting option for optical lithography of logic ICs or ASICs. Even though attPSM is not an aggressive or strong phase-shifting scheme, as has been demonstrated in Fig. 2.25, the economical factor alone has kept it alive all the way down to 65 nm node. As the photomask manufacturers pointed out, few IC manufacturers have required alternating PSMs. They are too expensive to make and the benefit out of that is not justifiable [53]. Process engineers can get by with attPSM and optical proximity correction (OPC) without using the alternating PSMs. The attPSM can be

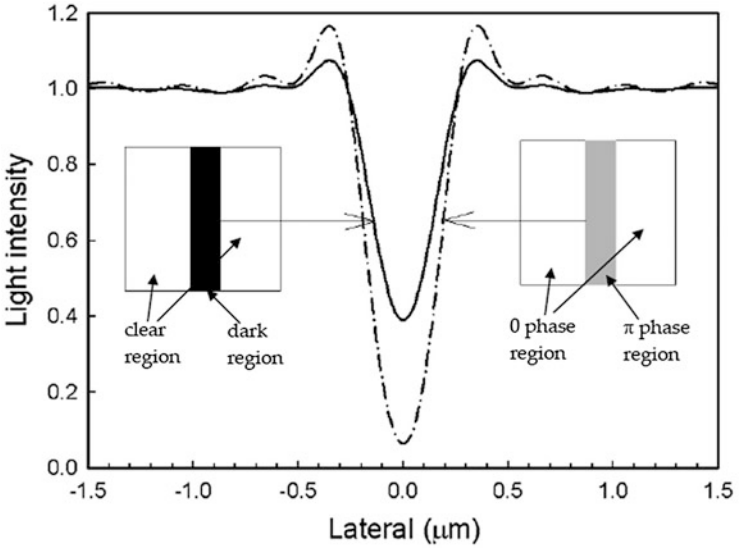


Fig. 2.26 Comparison of CPL mask (right) with conventional binary mask (left)

used not only in DUV but also in EUV lithography. Because the EUV masks are not refractive but reflective, the phase shift and attenuation are created by varying the absorber material and layer structure [54].

As mentioned earlier, alternating PSM is good for highly repetitive features such as the memory cell in DRAM-type ICs, but not for randomly distributed features which are common in logic-type ICs. To tackle this challenge, a variation of phase-shifting scheme shown in Fig. 2.24i–e, called chromeless phase lithography (CPL), was developed in the early 2000s [55]. Imagining that the transmission of phase-shifting layer on an attPSM becomes 100% instead of 5–15%, then at any edges of a mask feature, there will be dark lines because of the cancellation of light where 0° phase and 180° phase regions meet. If the two phase edges come reasonably close, then the two dark lines become merged into one dark line, which is equivalent to a dark line produced by a conventional binary mask (clear field mask) but with much higher contrast, because of the phase-shifting. This is illustrated in Fig. 2.26 where the aerial image of an opaque line from a binary mask is compared with that of an all-transparent mask but with a phase shift. The all-transparent mask produces the same dark line aerial image as the conventional chrome-on-glass mask but has much high image contrast, shown by the simulated light intensity in Fig. 2.26. As it is known that negative tone resist image in IC manufacturing is always realized by positive tone photoresist combined with a clear field mask, the CPL can well replace the conventional clear field mask to create such negative tone resist image with substantially improved resolution.

Dark aerial images from CPL are always formed by two phase edges in close proximity. The image formation is intrinsically symmetrical with no unwanted

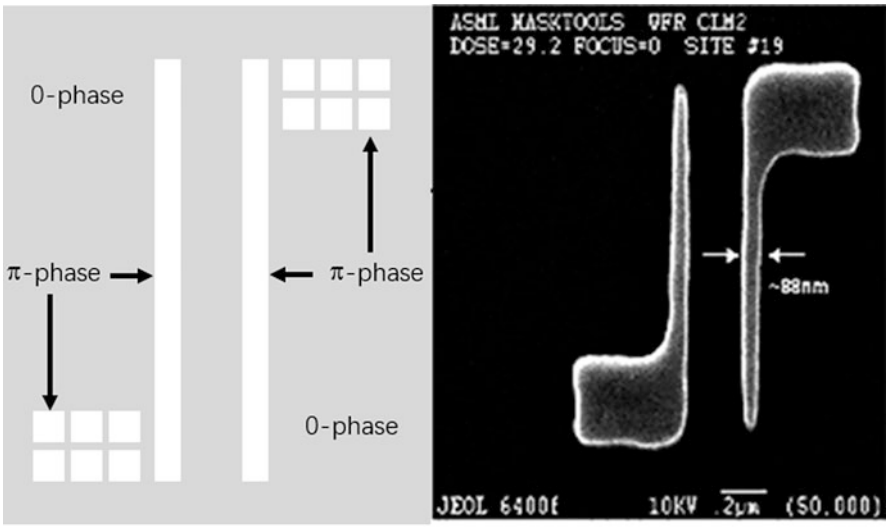


Fig. 2.27 (**a**) CPL mask and (**b**) printed resist patterns. (Reprint from Ref. [56] with permission)

phase-edge images. The CPL layout needed to form a resist pattern on the wafer is not substantially different from that of a binary chrome mask. CPL can be made by etching the quartz substrate of a conventional chrome-on-glass mask blank to create π-phase depth after the initial chrome patterning. The chrome is then removed in subsequent steps to reveal non-phase-shifting areas. Both "π-phase background areas" and "0-phase patterning areas" share the same quartz substrate. It looks like a conventional mask but without chrome patterns. By careful design of π-phase patterns, dark random patterns can be formed in photoresist. An example of CPL mask and its resist image are shown in Fig. 2.27 where irregular resist patterns are formed by phase-shifting features instead of chrome features. For design layouts with both large and small features on the same mask, a hybrid CPL mask can be used. A hybrid CPL mask contains both phase-shifting features and chrome features. For large patterns, chrome features can be used as it does not require high resolution. For small patterns, π-phase patterns can be used to improve image resolution. Because of the mixture of phase-shifting regions and chrome regions, two steps of processes are needed to fabricate the hybrid CPL mask. CPL performance can be further enhanced by high-NA imaging optics and off-axis illumination. In combination with OAI and 0.85NA optics, 70 nm features with minimum pitch of 160 nm have been produced at 193 nm wavelength illumination, which corresponds to k_1 factor of about 0.3 [56].

2.7.3 *Optical Proximity Correction (OPC)*

Because of diffraction, fine details of an object could be lost through an optical imaging process. As illustrated in Fig. 2.14b, the fine details of an image are contained in those high spatial frequencies of diffracted lights. The fidelity of an image can only be maintained if those diffractions of high spatial frequency can be imaged. Loss of fine details implies the image is no longer sharp but blurred. In optical lithography of IC patterns, this blur of mask pattern image is translated into corner rounding and line-end shortening and is called optical proximity effect. Such effect is demonstrated in Fig. 2.28a, where the simulated optical image is shown to compare with the original mask pattern. Line-end shortening and corner rounding

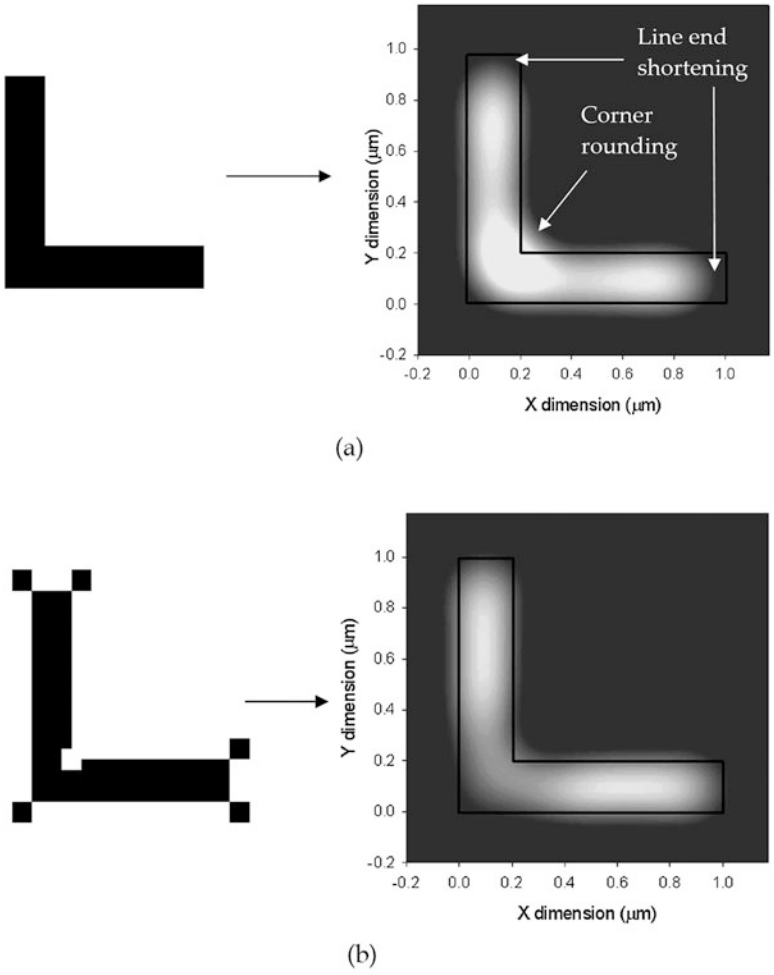


Fig. 2.28 (**a**) Optical proximity effect results in corner rounding and line-end shortening; (**b**) optical proximity correction modifies the mask design to restore the pattern fidelity

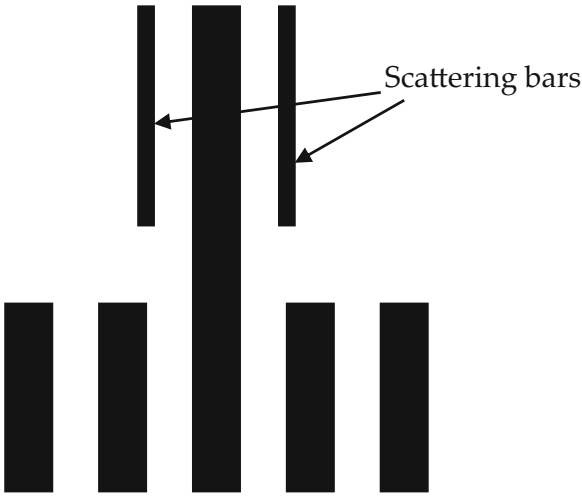


Fig. 2.29 Mixture of dense and isolated lines with added scattering bars

are apparent. Increasing numerical aperture (NA) and reducing illumination wavelength (λ) are able to reduce the proximity effect. However, as described in previous sections, there are limits as to how short wavelength and how high NA can be employed. Failing all that, the remedy is to modify the original design layout to compensate the line-end shortening and corner rounding. This is the idea of optical proximity correction (OPC) [57]. OPC is to deliberately distort the original design to balance the image intensity. For example, overexposure region can be reduced by taking away part of the pattern; underexposure region can be enhanced by adding some extra features. Figure 2.28b shows the design pattern after OPC and its intensity contours. Comparing with Fig. 2.28a, the OPC has made the image closer to its mask design. OPC can also be realized by modifying the transmission of different parts of a mask pattern using the so-called grayscale mask technique [58]. Although the OPC design looks different from the original design, those added features are all sub-resolution, meaning that they are not resolvable by the imaging optics and hence will not be printable in photoresist.

Another compensation technique is to use the so-called scattering bars, or sub-resolution assist features (SRAF) [59]. These are narrow lines or spaces placed adjacent to a primary feature in order to make a relatively isolated line behaving more like a dense line lithographically. For example, Fig. 2.29 is a group of lines. The lower half of the central line can be considered as dense line because it is closely adjacent to other lines. The upper half of the central line, however, is an isolated line. Upon optical exposure, the resultant CD (critical dimension) of the central line would be different in its upper section from its lower section. The upper section will be narrower than its lower section. This is the optical proximity effect. To make up the difference, one can either deliberately widen the upper section of the central line, which is called biasing, or add two narrow bars on both sides of the upper section, as shown in Fig. 2.29. These two narrow bars are called scattering bars or SRAF. They are too narrow to be imaged on photoresist (which is the reason why

they are called sub-resolution features), but the light passing through these bars can compensate the low intensity of light of the isolated line, so that the CDs of both dense lines and the isolated line are of the same width. SRAF is a complementary technique to the OPC methods mentioned earlier. SRAF is only workable if there are spaces nearby to allow placing them.

OPC has become a standard technique in sub-100 nm optical lithography. The implementation of OPC faces two challenges: one is the placing and routing of OPC features, and the other is the fabrication of OPC mask. OPC features are added after the IC design is done. It is a complicated task to determine where to place what type of OPC features or how to modify the original design. In the early days of employing OPC technique, placing the OPC features was based on some rules which were derived from experience and computer simulation of aerial images of some typical shapes [60]. Rule-based OPC has the advantage of fast correction because patterns are modified according to a set of rules. However, the accuracy is poor as the number of rules is limited, which cannot cover all the possible situations (variety of pattern shapes). The other method is the model-based method [61]. The execution of model-based OPC is to first flatten the hierarchy mask design layout into millions of segments; modelling is performed on each segment (1–2 μm range which is the distance of proximity interaction); the simulated pattern edge is compared with target pattern edge; difference called the edge placement error (EPE) is determined, and then the edge movement is performed to minimize the EPE. The final step is to resemble all the corrected segments back to the mask layout [62]. A post-OPC verification is also needed to check the validity of mask layout. In most cases, verification with a specific exposure tool and a specific wafer process is necessary. The model-based method is obviously more accurate but very time-consuming. With the advances in computer technology, simulation power has increased exponentially in recent years. Now the mainstream OPC methodologies are mostly model-based and different OPC software is readily available from commercial vendors.

All OPC features are sub-resolution, which means they are much smaller than the normal mask patterns and not imageable by the optical system. An OPC feature should be at most half the size of a normal feature. For 65 nm generation IC, the mask feature will be 265 nm for 4:1 reduction optical lithography. The OPC feature on mask should then be less than 130 nm. Making such small patterns on a mask plate can be considerably challenging. Currently there are two mainstream technologies for mask manufacturing, laser direct write and electron beam direct write. Laser tools are fast and have high throughput, while electron beam technique has the high resolution but slow. Figure 2.30 shows part of an OPC mask made by electron beam lithography, where the smallest features (serifs) are only 100 nm. Any distortions introduced during mask making can potentially degrade the OPC performance [63]. Other challenges are mask inspection and repair for these very small OPC features, and data handling of OPC design. The data file for a full mask OPC design at 90 nm technology is as large as 263GB and increases to 729GB at 65 nm technology [64]. With such a large data volume, not only computer processing time but also hard disk accessing time has become limiting factors to the mask making

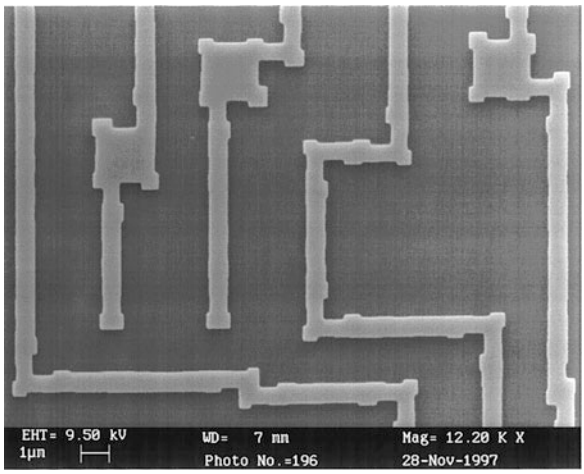


Fig. 2.30 SEM image of an OPC mask showing OPC features added to normal mask features

throughput. All have contributed to the escalating cost of masks for the sub-100 nm generation IC manufacturing.

The ultimate solution for achieving the best optical lithography result is to design the photomask by working out how an ideal image is generated and then reversing back to the "perfect" mask layout. This is the concept of "inverse lithography technology" (ILT) which was first proposed more than 30 years ago [65]. ILT is a rigorous mathematical approach to the design of photomasks for optimal lithography and can be viewed as an ultimate solution for OPC, or replacement of OPC, because the mask design has been optimized pixel by pixel to achieve the perfect balance of light intensity distribution on image plane. To reduce the difficulty in mask fabrication, the algorithm of finding the mask design can be tuned to different degree of aggressiveness. Less aggressive optimization can have a mask design which is easier to fabricate. Figure 2.31a shows two mask patterns from ILT design. The top one is an aggressive design and the bottom one is a nonaggressive design. The nonaggressive design lacks fine details; therefore, the mask is easier to make. Figure 2.31b shows the relevant resist images printed on wafer with the ILT masks.

There have been two obstacles hindering the adoption of ILT into mainstream IC manufacturing: (1) huge amount of computer power needed to process the mask data and (2) making curvilinear mask features which are preferable for the best imaging result. In recent years, two newly developed technologies provided solutions to overcome the obstacles: parallel graphic processing IC, i.e. graphic processing unit (GPU), and multi-e-beam writer [66]. GPU enabled fast calculation of imaging data, and multi-e-beam lithography (see Sect. 3.7.2 of Chap. 3) allowed fast manufacturing of complicated ILT masks with curvilinear features. These technologies enabled full-chip ILT for the first time in 2019 [67, 68]. The optimizations from illumination to mask, including PSM, OPC, and ultimately ILT, are all based on computer simulations. The fields have been unified under the term of "computational lithography" [69], which has become an integral part of optical lithography

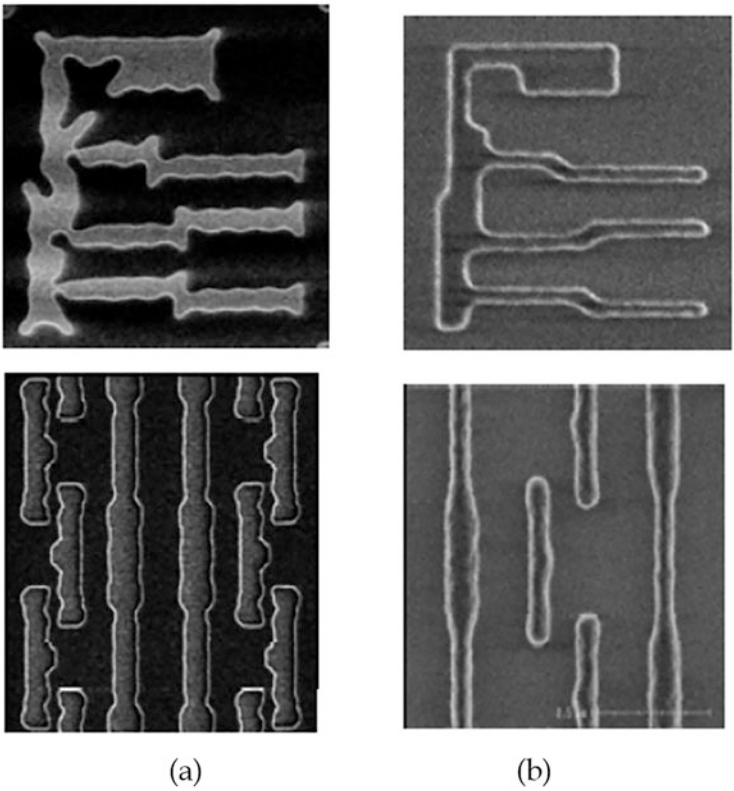


Fig. 2.31 (**a**) Aggressive (top) and nonaggressive (bottom) ILT masks. (**b**) Resist images printed from aggressive design (top) and nonaggressive design (bottom). (Reprint from Ref. [72] with permission)

process. Without computational lithography, it would be impossible for chipmakers to manufacture the latest technology nodes. Considerable effort and cost have been saved by doing computational optimization first before going into experimental trials. A survey conducted in 2022 indicated that more than 90% of respondents said they have used ILT in some critical layers of masks [70]. Adoption of machine learning technology can further enhance the efficiency of computational lithography [71].

2.7.4 Photoresists

All the techniques described in previous sections are merely forming a light image on the surface of a photoresist layer. The mask pattern that represents an IC design will have to transform into a photoresist pattern so that it can be further transferred

into the substrate material. The description of basic techniques for processing photoresists, such as exposure, baking, development, etc., can be found in many literatures [3, 73, 74]. Section 2.3 of this chapter has also given a general overview of the basics of photoresists and lithography process. Although the basic process steps are more or less the same whether for patterning of sub-micrometer features or for patterning of sub-100 nm features, more technical difficulties and challenges are encountered when the photoresist patterns are less than 100 nm. The four key parameters for photoresists and their processes at sub-100 nm resolution regime are sensitivity, contrast, line edge roughness (LER), and etch resistance. The etch resistance will be discussed in Chap. 7. This section will focus on sensitivity, contrast, and line edge/linewidth roughness.

2.7.4.1 Sensitivity

Since the optical lithography moved to deep UV wavelength at 248 nm, chemically amplified (CA) resists, which are distinctively of high sensitivity, have become the working resists. The CA resist technology was pioneered by IBM in the 1980s in order to accommodate early deep UV light sources [75]. At the time, the choice of exposure light sources was still mercury lamps which were widely used for g-line and i-line optical lithography. Although excimer lasers emerged as new sources of deep UV wavelength radiation, they were in a very early development phase, and there was concern over their ability to operate reliably. Mercury lamps, already proven in manufacturing, were deemed a more reasonable approach. However, the deep UV spectrum in a mercury lamp was rather weak and hence required very fast resists, leading to the development of CA resists.

The high sensitivity characteristics of CA resists rely on a substance called "photoacid generator" (PAG). In these resists, the lithographically important chemistry is not driven directly by photons absorbed during the exposure step. Instead, an acid formed upon exposure of a PAG catalyzes chemical changes in the resist during the thermal baking step, which is called post-exposure baking (PEB). Since the reaction is catalytic, the acid is regenerated after each chemical reaction; the same acid molecule can participate in further reactions, hence named chemical amplification. Each PAG can induce 800–1200 more photoacid molecules after the PEB step [76]. These photoacid molecules convert the exposed region of the CA resist into soluble or insoluble, depending whether it is a positive or a negative resist. Because only a small amount of photon energy is needed to generate the initial photoacid "seeds" and the major chemical process occurs during PEB, they are therefore of very high sensitivity. While conventional i-line resists need more than 100 mJ/cm^2 photon energy, CA resists need only 10–20 mJ/cm^2 photon energy.

The very high catalytic chain length in CA resists, which provides the high sensitivity, is also problematic because the generated photoacid molecules can be deactivated by airborne amine diffusion into the resist, which results in "T-topping" or swelling in resist pattern. Amine concentrations as small as 5 ppb in air are enough to induce the T-topping, depending on the length of time the resist is exposed to

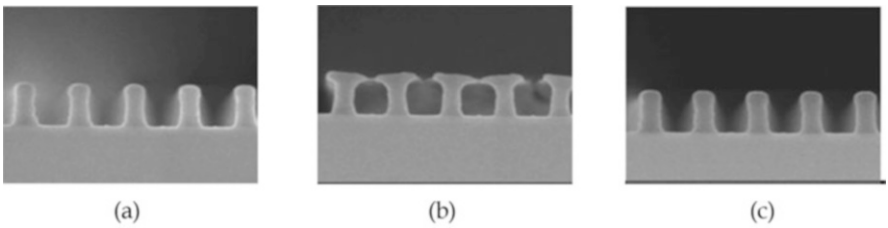


Fig. 2.32 SEM images of CA patterns (**a**) without delay, (**b**) with 45 min delay in clean room ambient, (**c**) with 45 min delay but kept inside exposure tool (150 nm CD and ArF lithography). (Reprint from Ref. [77] with permission)

ambient air having trace of amine. For example, Fig. 2.32 compares the CA resist patterns when the exposed resist was developed immediately without delay (Fig. 2.32a) and when there was a delay of 45 min in clean room ambient before development (Fig. 2.32b). T-topping is formed because of the exposure to ambient air. When the wafer was delayed for 45 min but kept in the exposure tool where the environment is enclosed and well filtrated, no T-topping was formed as shown in Fig. 2.32c. Improvement in resist chemistry to make it less sensitive to amines is a possible solution. But the most straightforward approach is to control airborne contamination with robust filters in the exposure tool or in the clean room. Chemical filters, in some cases, have moved from traditional carbon-based media to new formulations, including highly porous polymeric materials that can provide improved removal efficiency and capacity. High rate of chemical amplification during the PEB step also increases process sensitivity to resist baking cycles, necessitating extremely tight control of baking temperatures and times. From the process control point of view, too high sensitivity of a CA resist may not be beneficial. A compromise in sensitivity could render a CA resist more stable in a production process environment.

Coming to the EUV era, the chemically amplified resists (CARs) used in DUV lithography are no longer suitable because they are more transparent at 13 nm wavelength than at 193 nm wavelength. A series of non-CA resists containing metallic elements have been developed. Details of EUV resists have been described in Sect. 2.4.2.4.

2.7.4.2 Contrast

High contrast is the key to achieve high resolution in a photoresist. There are two different contrasts: one is the contrast of optical imaging, and the other is the contrast of photoresist, each of which has different definition [3]. However, one can tell the high contrast optical imaging or high contrast resist structure simply by comparing the aerial images (Fig. 2.33) or the resist profiles (Fig. 2.34a, b). It must be pointed out that high contrast of optical imaging does not necessarily lead to high contrast

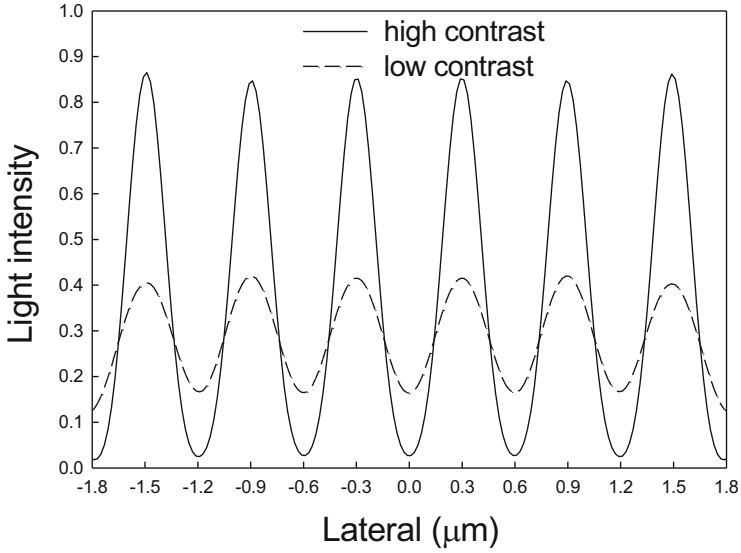


Fig. 2.33 Low and high contrast optical imaging

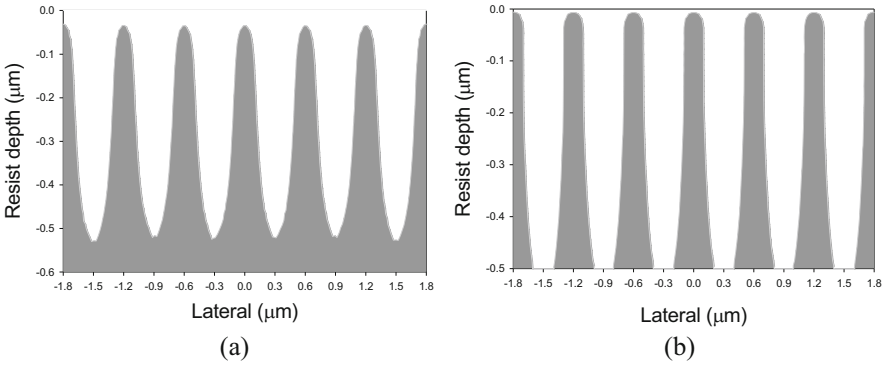


Fig. 2.34 (**a**) Low contrast resist pattern, (**b**) high contrast resist pattern

resist profile, whereas a high contrast photoresist can, to some extent, compensate a low contrast optical imaging to produce high resolution resist patterns. That is why the photoresist is a vital part of resolution enhancement technology.

To achieve high contrast, the resist material should be less absorptive to UV radiation. With low UV absorption, it is easier to achieve uniform distribution of photon energy through the depth of photoresist layer. Too absorptive will result in more photon energy being absorbed at the top and less at the bottom. As a consequence, the development of resist will be fast at the top and slow down through the depth, resulting in slopped pattern profile instead of a sharp and vertical resist profile. Unfortunately, materials are always more absorptive at shorter wavelength

radiation. Unlike i-line resists which work at 1 μm thickness, CA resists for deep UV have to work at less than 0.5 μm thickness in order to compensate for the high absorption of photons. Alternatively, a bilayer resist system can be used, with a thin layer of resist on top as the imaging layer [78].

2.7.4.3 Line Edge/Linewidth Roughness (LER/LWR)

For very small feature dimension, the edge roughness of developed resist lines can be a significant fraction of the linewidth. A measure of the line edge roughness (LER) is the 3 sigma standard deviation of the actual line edge relative to an average line edge. The roughness can be quantitatively measured by a scanning electron microscope (SEM) or an atomic force microscope (AFM). Figure 2.35 shows some CA resist lines with average linewidth of 111 nm. The roughness, however, is 37 nm as measured by SEM. The LER for CA resists at sub-100 nm feature dimension has become a serious issue. LER was included in the International Technology Roadmap for Semiconductors (ITRS) since 2001, as one of the key requirements on photoresists. ITRS indicates that the LER should be below 5% of the critical dimension (CD). For the 90 nm IC technology node, the LER should be less than 4.5 nm. While the functional lines patterned for ICs have shrunk below 22 nm, LER did not shrink accordingly. This is particularly true for CA resists.

It is still unclear what the causes of LER are. A number of statistically fluctuating effects at these small dimensions may contribute to the LER, such as shot noise (photon flux variations), statistical distributions of chemical species in the resist (such as photoacid generators), the random walk nature of acid diffusion during chemical amplification, and the nonzero size of resist polymers being dissolved during development [79]. Careful measurements and evaluation of LER data have ruled out the correlation between polymer molecule size and LER [80]. There have been attempts of using statistical modelling to explain the LER measurement data. The general trend is that the higher the resist sensitivity, the rougher the resulting line

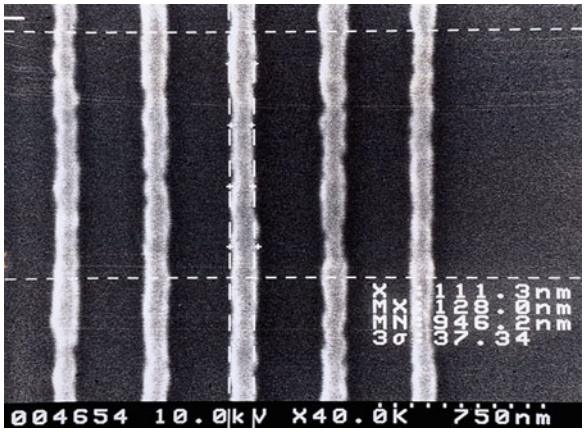


Fig. 2.35 SEM image of CA resist lines with average linewidth of 111 nm and line edge roughness of 37 nm

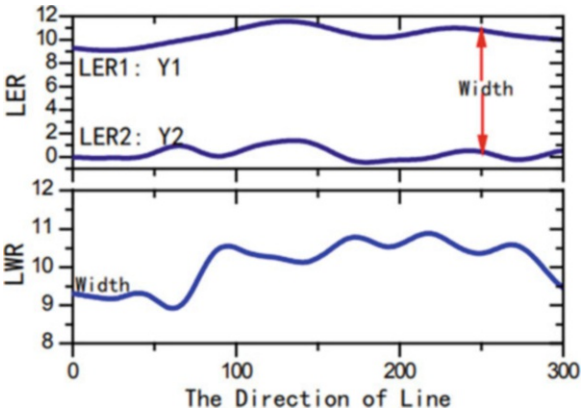


Fig. 2.36 Schematic illustration of linewidth roughness (LWR) caused by line edge roughness (LER). (Reprinted from Ref. [82] with permission)

edge. CA resists have not been able to meet the requirements of both high sensitivity and low LER. By analytical modelling of three key resist processing steps, the exposure, PEB, and development, it was demonstrated that the observed trade-off between resist sensitivity and LER is fundamental, which implies the basic limitation of achieving simultaneously low roughness, low dose, and high resolution in any standard CA resist process [81].

At the feature dimension below 16 nm, LER can lead to linewidth variation. A new term called linewidth roughness (LWR) was adopted to reflect the seriousness of problem. Figure 2.36 schematically illustrates the LWR. At the nominal linewidth of 10, the LER results in the variation of linewidth from 9 to 11. Correlations between the LER and LWR exist statistically [82], and the LWR can definitely lead to the variation of device property [83]. The current IC manufacturing requires the resist to satisfy LWR $<15\%$, in addition to resolution <10 nm and exposure dose as low as 20 mJ/cm^2 [84]. Fortunately, the LER became less pronounced in EUV lithography, because of using non-CA resists with inorganic components [85].

In addition to the aforementioned properties, there are a few other factors that can critically influence the production worthiness of 193 nm resists. They are the linearity of translating mask pattern to wafer, line slimming due to electron radiation during SEM inspection, and pattern collapse during wet development. They are briefly described in the following.

Usually, the mask features are linearly imaged onto a wafer. A variation of mask CD will cause variation of wafer CD, except 4 times smaller if it is a 4:1 reduction optics. However, at sub-100 nm resolution regime with low k_1 factor, small variations of CD on the mask can cause disproportionately large changes in wafer CDs. This amplification of mask errors is called the mask error enhancement factor (MEEF). For a linear imaging, MEEF $= 1$. If MEEF >1, the errors in mask CDs are nonlinearly transferred to the wafer even if the mask CD itself gets geometrically reduced by the proper ratio (4:1). For the dense line/space pattern, both theory and experiment indicate that mask errors are amplified by 2.5 times at $k_1 = 0.35$ [86]. Fundamentally, the nonlinear imaging originates from the decrease of aerial

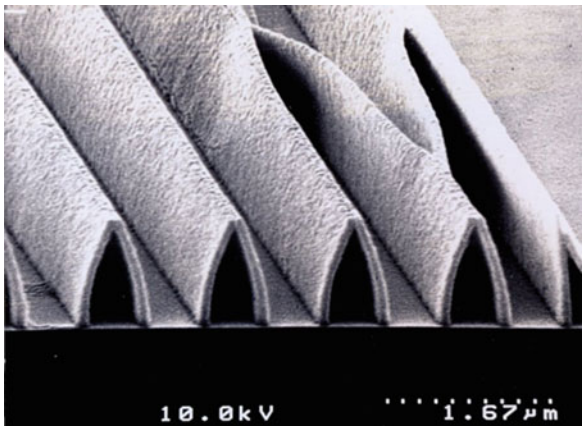


Fig. 2.37 Resist pattern collapse

image contrast because of the loss of high spatial frequencies in projection imaging. An easy solution to counter the large MEEF can be simply biasing the mask feature. In the meantime, mask errors and defects have to be reduced at low k_1 optical lithography, to avoid the detrimental effect of MEEF.

With IC manufacturing entering sub-100 nm regime, wafer inspections have widely adopted CD-SEM, a scanning electron microscope-based CD inspection and measurement, because of its much higher resolution capability than optical inspection methods. However, it has been observed that patterns of 193 nm resist shrink under electron irradiation during CD-SEM inspection, although very low electron beam energy (below 1 keV) and beam current (below 10pA) are used in CD-SEM. The shrinkage due to electron irradiation ranges from 0.2 nm to 0.6 nm per measurement scan for different groups of 193 nm resists [87]. The longer the electron beam dwells on the resist pattern, the more significant the feature shrinkage becomes. Among the two common polymer systems of 193 nm CA resists, acrylate and methacrylate, the methacrylate system seems to suffer more severe shrinkages. After 30 scan measurements at 800 eV electron energy, the methacrylate-based resist shrinks up to 30 nm for contact hole patterns [88]. The exact mechanism of e-beam-induced shrinkage is not clear. While the 193 nm resists will still be improved, a reasonable approach to this problem is to readjust the CD measurement procedure to minimize resist deformation.

Collapse of high aspect ratio resist patterns is a well-known phenomenon. After development and rinsing, capillary forces are present between closely packed resist structures. Unbalanced capillary forces lead to a net force on the pattern during drying process. Pattern collapse occurs if the capillary forces exceed the elastic restoring forces of the polymer [89]. Figure 2.37 demonstrates the typical resist pattern collapse phenomenon. A figure of merit for the collapse behavior is the critical aspect ratio of collapse (CARC), which is defined as the aspect ratio for a given width and pitch with 0% of line collapse [90]. The worst line collapsing happens for asymmetric structures consisting of an asymmetric pitch. On the other hand, pattern collapse occurs only for unbalanced forces. Therefore, the collapsing

effect is more probable in the middle part of a line than at their end, because water level is easier to equalize at the line end. For the same reason, the probability of line collapse is higher for longer lines in comparison to shorter lines. Because the capillary force is proportional to the surface tension of the rinsing liquid, using a supercritical fluid, such as liquid CO_2, instead of water for rinsing can eliminate the capillary force effect [91]. The simplest approach to reducing pattern collapse is to reduce the resist film thickness. However, for single layer 193 nm resists, the lower limit for the resist thickness has to be in the range of 300–500 nm because of the etch resistance consideration. In this case, using multilayer resist systems can be an option.

2.7.5 Multi-patterning

Before the EUV lithography was ready for production, the IC industry was left with the only choice of 193 nm immersion lithography (193i). All the aforementioned RETs were employed, which helped to reduce the k_1 factor. However, these were not enough for the IC industry to keep pace with the demand for further reduction of circuit feature dimensions. Starting from 20 nm node downward, the circuit features became too dense on the mask, making it more difficult to pattern discernible features on the wafer. Particularly when OPC and sub-resolution assist features (SRAF) were used, as seen in Fig. 2.38 [43], there were simply not enough spaces to place these OPC features. On the other hand, isolated features or less dense features are much easier to pattern, as has been illustrated in Fig. 2.21. The solution to the dense feature problem is to split the circuit layout into two separate masks and image them in two patterning process, the so-called double patterning (DP).

There are three variations of double patterning methods: litho-freeze-litho-etch (LFLE); litho-etch-litho-etch (LELE); and self-aligned double patterning (SADP). Figure 2.39 schematically compares the LFLE and LELE with single patterning (Fig. 2.39a). In the LFLE patterning process (Fig. 2.39b), the exposed and developed resist structures are chemically modified (freeze) before applying the second layer

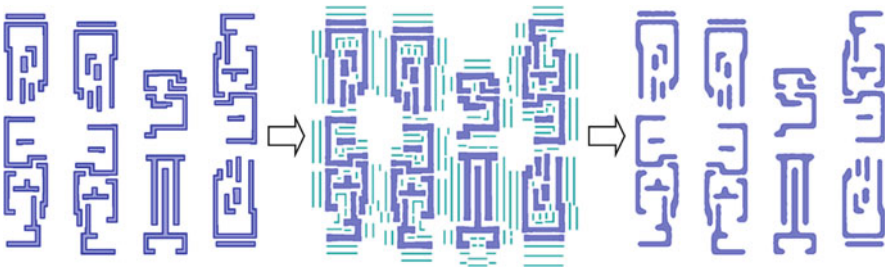


Fig. 2.38 Example of original mask layout to OPC/SRAF-added mask layout to final resist image. (Courtesy of David Abercrombie, Semiconductor Engineering [43])

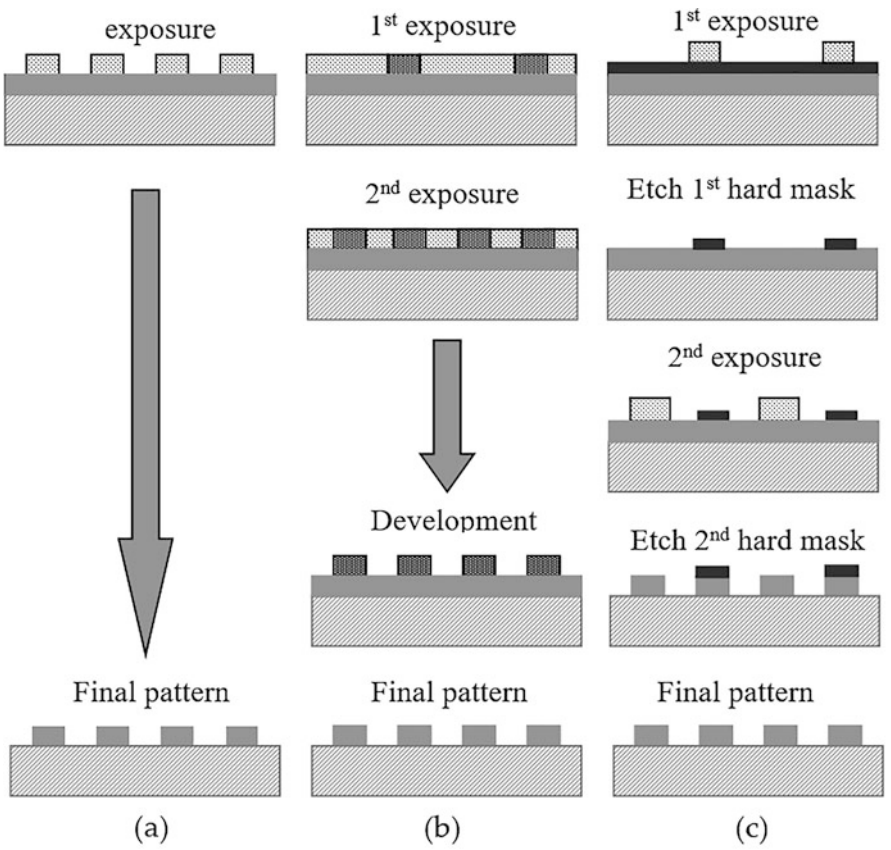


Fig. 2.39 Comparison of (**a**) single patterning with (**b**) LFLE and (**c**) LELE

resist and exposure. The "freeze" makes the resist no longer respond to the second exposure. The combination of both exposures produces dense resist patterns. In the LELE patterning process (Fig. 2.39c), two layers of hard masks are used. Each hard mask is patterned by photoresist exposure separately. The composition of both masking patterns becomes the final dense masking patterns. Both LFLE and LELE double the pattern density. However, both require two sets of masks and two exposures, which have to be accurately aligned to each other within 1.5–3 nm.

While the LFLE and LELE require very stringent alignment accuracy for the two-step exposure, self-aligned double patterning (SADP) is much simpler. As illustrated in Fig. 2.40, SADP only involves single exposure. The doubling of pattern density is realized by spacers. As the deposition of spacer material is naturally conformable to the resist pattern, no alignment is required. The directional plasma etching can remove the top and bottom spacer material to create spacer patterns which double the density of original resist pattern. The spacer then acts as mask to

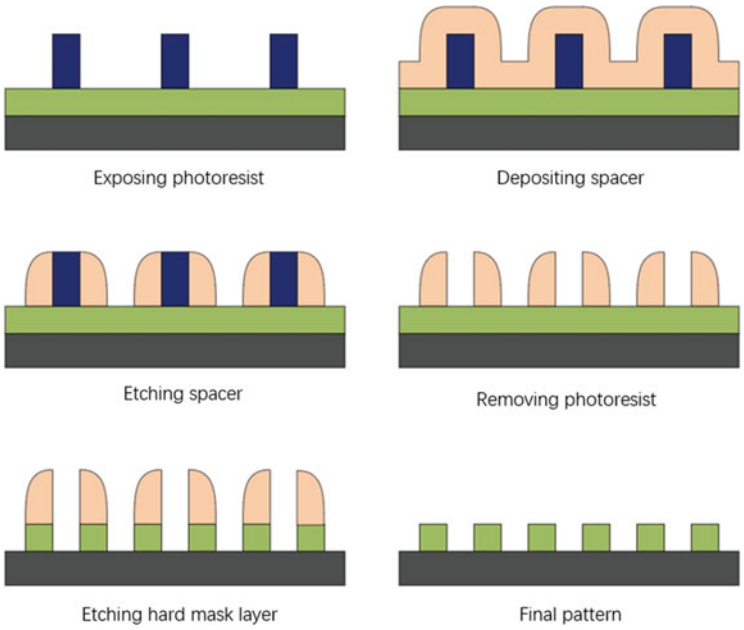


Fig. 2.40 Self-aligned double patterning (SADP) process. (Courtesy of Cepheiden's artwork from Wikimedia Commons)

pattern the hard mask layer. Finally, the spacer is removed by isotropic etching. Section 9.2 of Chap. 9 will further elaborate this type of indirect nanofabrication. Although the SADP is simple to implement and less process steps are involved, it is only suitable for line/space type of pattern. For nonperiodic pattern such as logic ICs, the patterns created by spacer may not be the patterns needed in real circuits. In practice, all the above three double patterning methods are being used.

Double patterning enabled the 193i lithography to go down 20/16/14 nm IC technology nodes. For 10 nm, 7 nm, or 5 nm nodes, triple patterning (TP) or quadrupole patterning (QP) has to be used. By repeatedly using the spacer concept eight times (octuple patterning), 6.25 nm half-pitch line/space patterns were created by 193i lithography [92]. The multi-patterning has truly broken the theoretical limit of k_1 factor, and even with EUV lithography, the multi-patterning will be needed when IC processing comes down to sub-5 nm regime (Fig. 2.19).

2.7.6 Design-Technology Co-optimization (DTCO)

As described in previous sections, to lower the k_1 factor, many RETs have to be used. The original mask layout for an IC design has changed considerably for the $k_1 < 0.3$ regime, after applying PSM, OPC, SRAF, and ILT, which can be vastly different

from the printed photoresist image on a wafer. In the earlier above 100 nm process technologies, IC designers do not have to know the lithography processes. They only need to follow a set of design rules which defines how small a feature should be and how close two features to each other should be, etc. Since IC manufacturing entered 90 nm node and below, the traditional implementation of more sophisticated RETs has seen design rules exploding. For example, the number of rules at 90 nm node rose four to five times, many of which are never explained to a designer whether they are related to lithography, etch, CMP, other manufacturing needs, or all of the above.

On the other hand, complicated RET measures and addition of RET features to the original design have put considerable pressure on IC process engineers. Some of the RET modifications may be sound from the optimization of optical imaging, but may be completely unmanufacturable, or with very low manufacturing yield. At 90 nm, IC manufacturers find that only less than half of IC designs operate as expected and more than half of them need a complete mask "re-spin" for acceptable yield and performance. With the design costs approaching $10 million or more and mask costs alone nearly $1 million, Design for Manufacturability (DFM) has become a key issue, and IC designers must consider manufacturability issue throughout the design cycle [93].

The methodology of DFM is to start with identification of manufacturability and then goes back to designers to tune the design in order to achieve the maximum manufacturing yield. One of the DFM approaches favored by IC manufacturers is to predict, during the design stage by model-based computer simulation, how the ideal design layout will print as contours in silicon. These contours are equivalent to 3D structure profiles produced through the flow of manufacturing processes, which are accurate representation of the final manufacturing results. From that, designers can predict the "hot spots" where the circuits tend to fail. Once the hot spots are identified, the design layouts will be manipulated to make sure those potential failure spots are eliminated [94]. This model-based prediction is different from previously described OPC simulation. It has incorporated not only optical imaging but also manufacturing parameters. A new set of lithography-friendly design rules can be established by the manufacturing process-related computer simulation [95, 96].

While the DFM is the design community responding to the manufacturing community, the introduction of double patterning below 20 nm node puts considerable constrains on designing process, and the optimization is not only required at the designer side but also required at the manufacturer side. Design-technology co-optimization (DTCO) became the norm [97]. DTCO encompassed DFM. However, the optimization at the manufacturer side is not only on the existing process technology but also on the choice of different technologies and different IC architectures, with the end goal to improve performance, power efficiency, transistor density, and cost. It should be pointed out that the situation became relaxed since the introduction of EUV lithography which led to the k_1 factor to jump to as high as 0.56. The EUV lithography enabled a different meaning of DFM (Design Friendly Manufacturing) [98]. However, the luck will soon run out as multi-patterning will be needed even for EUV lithography at sub-5 nm node, as indicated in Fig. 2.19.

2.8 Near-Field Optical Lithography

Projection optical lithography has been well developed over the course of more than 40 years. It has become the mainstream lithography technology for high-volume IC manufacturing. However, one should not forget that optical lithography was originated from contact printing. Contact printing relies on a mask aligner to print a mask pattern onto a photoresist in 1:1 ratio. Although mask aligner is no longer used in modern IC manufacturing for many reasons such as low throughput, no image reduction, and easy damage to mask plate, it still has a role to play in large feature size lithography, particularly in manufacturing of MEMS (microelectromechanical system) devices. In recent years, contact optical lithography has been revitalized as an alternative low-cost technique for sub-100 nm patterning technique.

It is known that projection optical lithography is a far-field optical imaging process, which has a fundamental resolution limit of $\lambda/2$ where λ is the wavelength. Contact optical lithography is by nature a near-field optical imaging process. There is effectively no space for light wave to travel between the mask opening and photoresist layer, except within the photoresist layer. At such an intimate contact, light waves no longer propagate as sinusoidal wave but as evanescent wave, which decay exponentially with distance. Therefore, near-field optical lithography is also called evanescent near-field optical lithography (ENFOL) [99]. The evanescent wave no longer behaves as sinusoidal wave which diffracts through small openings, and the imaging resolution is diffraction limited. The resolution of near-field optical lithography is only limited by the mask opening, not by optical wavelength. That is the reason that near-field optical lithography can achieve sub-100 nm resolution. One only needs to make the mask features small enough.

Near-field optical lithography evolved from the contact printing but is not the same as a conventional contact printing. Instead of using a hard glass mask, a thin conformable mask is used. Figure 2.41 shows the setup of near-field optical lithography, where the mask is made of 2-μm-thick silicon nitride membrane with chrome patterns. The membrane mask is supported by a silicon frame and mounted onto a vacuum chuck. The intimate and conformable contact is realized by vacuum between the membrane mask and photoresist-coated substrate. Though not a transparent mask, the thin nitride membrane can transmit 61% of g-line wavelength (436 nm) and 29% of h-line wavelength (405 nm) irradiation from a broadband mercury arc lamp [100], which is adequate to expose commercial g-line photoresists.

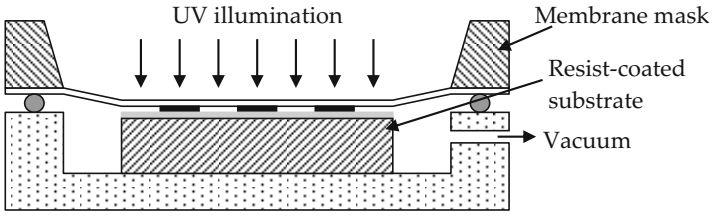


Fig. 2.41 Setup of near-field optical lithography. (Reprint from Ref. [100] with permission)

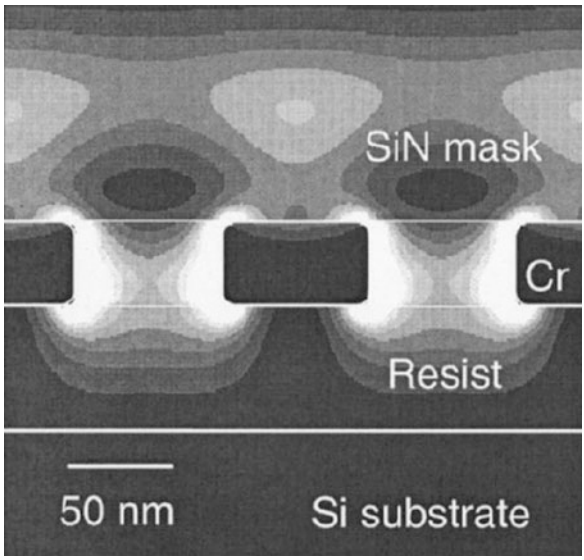


Fig. 2.42 Computer-simulated contours of squared electric-field intensity $|E|^2/|E_0|^2$ where $|E_0|^2$ is the incident intensity. (Reprint from Ref. [100] with permission)

Because the evanescent wave can only travel very short distance, the ENFOL can only be performed on very thin photoresist, normally less than 100 nm thickness. Such a characteristic is best explained by the computer simulation of electromagnetic field distribution, as shown in Fig. 2.42. It shows contour plots of the squared electric-field intensity for a cross section through a 140-nm-period and 40-nm-thick Cr mask illuminated from above with 436 nm light. The exposure is into a 60-nm-thick photoresist layer on silicon. The black and white regions correspond to the ratios of $|E|^2/|E_0|^2$ from <0.1 to >1.3, where $|E_0|^2$ is the incident intensity.

Two properties are revealed for the evanescent wave propagation. The first is that the penetration of evanescent wave is only about the top 40 nm of the resist layer, as seen in the bright region from the Cr mask opening into the resist layer, which implies that complete resist clearance of exposed patterns will be difficult to achieve. In this case, a short O_2 plasma etching will have to be performed to remove this residual resist layer prior to transferring the resist pattern into the underlying silicon. The second is the large field enhancements seen at the edges of the mask pattern made of metal conductor, where the bright regions represent $|E|^2/|E_0|^2 > 1$, indicating that the local electrical field intensity can be greater than the incident intensity. The local enhancement is strongly dependent on the sharpness of the edges, which is a possible explanation that patterns produced by evanescent near-field lithography are always of high line edge roughness.

The fact that electrical field of incident light has been enhanced or amplified indicates there is something else happening in the evanescent near-field imaging process. This is the surface plasmon polariton (SPP) phenomena. SPP is a charge density wave on the surface of material which has plenty of free electrons, such as silver or gold, with silver being the best material for exciting the SPP. It is an

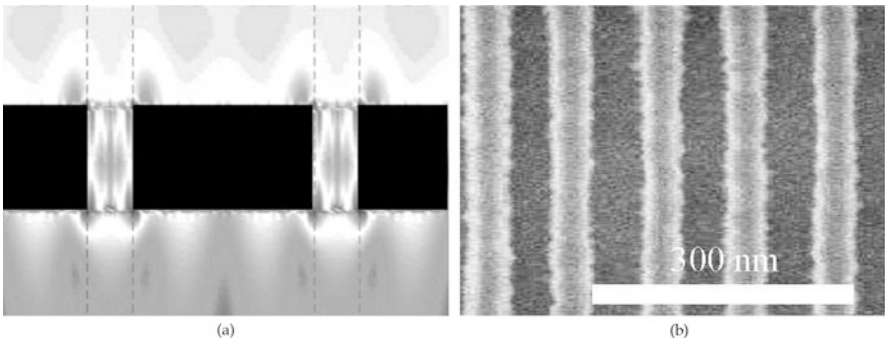


Fig. 2.43 (**a**) Computer simulation of near-field distribution, (**b**) experimental lithography result. (Reprint from Ref. [103] with permission)

interaction between the electromagnetic field of incident light wave and the conducting free electrons on the surface of metal thin film or nanostructures [101]. The SPP wavelength depends on the wavelength of incident light and the permittivity of conducting material, which can be much shorter than the incident light wavelength [102].

Since the SPP wave can propagate along the interface of the conducting material and a dielectric material, interference can occur, resulting in standing waves with periodically distributed maxima and minima. Figure 2.43a shows the computer-simulated near-field distribution for two well-separated mask opening aperture illuminated by g-line ($\lambda = 436$ nm) light wave [103]. The mask material is silver. The separation of the two apertures is 300 nm. The simulation reveals clearly that the electrical field maxima exist not only at the mask apertures but also between the apertures where there are no openings to the incident light. This is the result of interference of SPP waves from both sides of the aperture. The computer simulation has been proved by experiment with exposure using a g-line illumination source and 50-nm-thick photoresist layer. A vacuum chuck is used to pull the mask onto the resist-coated substrate to ensure intimate contact. Figure 2.43b shows the exposed 50 nm resist line patterns. Although the pitch of line grating on mask is 300 nm, the actual photoresist pattern has a pitch of only 100 nm. The two extra lines between the mask features are created by the interference standing wave of SPP. It is noticed that the resist feature size is only 1/10 of the exposure wavelength. Similar SPP interference can also happen within the opening of a mask feature, which creates smaller features in photoresist than the mask feature itself [104]. It is also noticed that the LER (line edge roughness) is quite significant. This is caused by local enhancement of electrical field.

Because photoresist can be exposed by SPP, not necessarily by incident light, then it is not necessary to have the mask in intimate contact with photoresist for patterning. A pattern can be printed onto photoresist, as long as there is a mechanism to generate the SPP pattern. This is the concept of silver superlens [105]. An example of superlens structure is illustrated in Fig. 2.44. A conventional photomask (not the

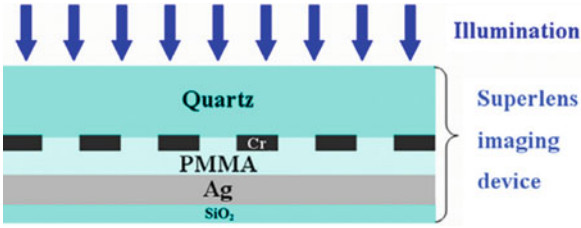


Fig. 2.44 Schematic of SPP superlens structure. (Reprinted from [106] with permission)

membrane-type conformable mask) is coated with 30-nm-thick polymethylmethacrylate (PMMA) as a spacer; then 50-nm-thick silver is coated as the superlens, as well as 30-nm-thick SiO_2 as a protective layer. Although light is not directly shining on a photoresist layer, exposure of the photoresist can still happen by the SPP radiation excited in the silver layer with the photomask. With this superlens, 110 nm line pattern has been transferred to photoresist with an i-line illumination [106]. Along this direction, plasmonic reflective lens and plasmonic cavity lens have also been developed for near-field optical lithography [107].

As described above, near-field optical lithography can indeed realize sub-100 nm patterning of photoresist using much longer illumination wavelength. However, there are a number of inherent drawbacks in this technology:

- *Intimate contact is required*. Because the near-field optical lithography relies on evanescent wave which decays exponentially with distance, the mask has to be intimately in contact with photoresist. This is the reason that a membrane-type conformable mask has to be used. A membrane mask is fragile. The requirement of conformable contact with photoresist-coated substrate makes it even more vulnerable to break. Membrane mask is also difficult to be made into large size. The superlens approach can alleviate the awkwardness of using a membrane mask. It still requires the metallic superlens layer to be in intimate contact with the photomask, which makes the reusability of patterned photomask greatly reduced. Some research has demonstrated that with the help of plasmonic cavity lens and off-axis illumination (OAI as introduced in Sect. 2.7.1), up to 80 nm air gap between the plasmonic lens and photoresist layer can be allowed [108].
- *Exposure depth is very limited*. Both evanescent wave and SPP wave decay very rapidly in photoresist, resulting in a very shallow exposure depth. The thickness of photoresist for near-field optical lithography should be no more than 100 nm. Such a thin photoresist layer makes subsequent pattern transfer by etching very difficult. Multilayer resist structure is commonly used, whereby the near-field optical lithography only patterns the top imaging layer and the resist pattern is transferred to underlayer by reactive ion etching [106].
- *LER is significant*. Because the charge density induced in the metal mask by incident light is highly concentrated at any sharp edges of a mask feature opening, the locally enhanced intensity can cause significant line edge roughness (LER) in the exposed photoresist features. There are two approaches that can reduce the

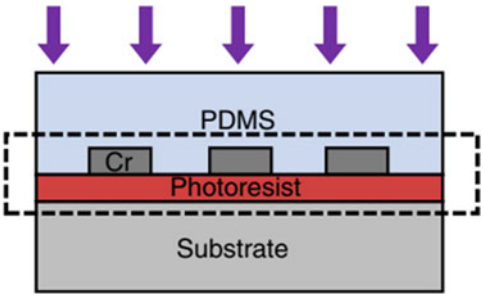


Fig. 2.45 Soft PDMS mask for near-field optical lithography [110]

LER: one is to improve the mask making process to reduce the LER of mask features; the second is to improve the photoresist material to make it less sensitive to the speckles of image intensity.

Nevertheless, near-field optical lithography is a simple and low-cost technology for making sub-100 nm patterns. There have been attempts to extend the technology for high-throughput patterning. Reported approaches include employing bowtie antenna (see Sect. 11.4.3 of Chap. 11) apertures as plasmonic lens array [109], or scanning an optical fiber probe in near field (see Sect. 5.3.2 of Chap. 5). Even if it may never become viable in mass manufacturing of ICs, this does not stop the technology to be used in laboratory for research wherever there are requirements for making sub-100 nm structures. Its similarity to conventional contact printing using a mask aligner makes the technology much more accessible than the state-of-the-art optical projection lithography.

The key to realize high resolution by near-field lithography is the near field which requires the mask in intimate contact with substrate. None of the aforementioned masks are able to be used repeatedly. There is a simple way to realize repeatable intimate contact between mask and substrate, which is to use a soft mask. The typical material for making soft mask is polydimethylsiloxane (PDMS). The softness of PDMS enables the intimate contact with substrate, as illustrated in Fig. 2.45 [110]. Since the PDMS is soft and elastic, it can be rolled onto a glass tube. By installing a UV light inside the glass tube, it becomes a rolling mask to carry out near-field optical lithography in a roll-to-roll manner [111]. A company named Rolith, which was acquired by Canadian company Metamaterial Technologies Inc. in 2016, commercialized the technology and built industrial-scale lithography systems.

2.9 Talbot Optical Lithography

There is an interesting optical phenomenon which was discovered by Henry Fox Talbot almost 200 years ago. He found that when a monochromic coherent light is shining through a mask with periodical openings, the images of mask openings reappeared at different distances from the mask and at some distance these images

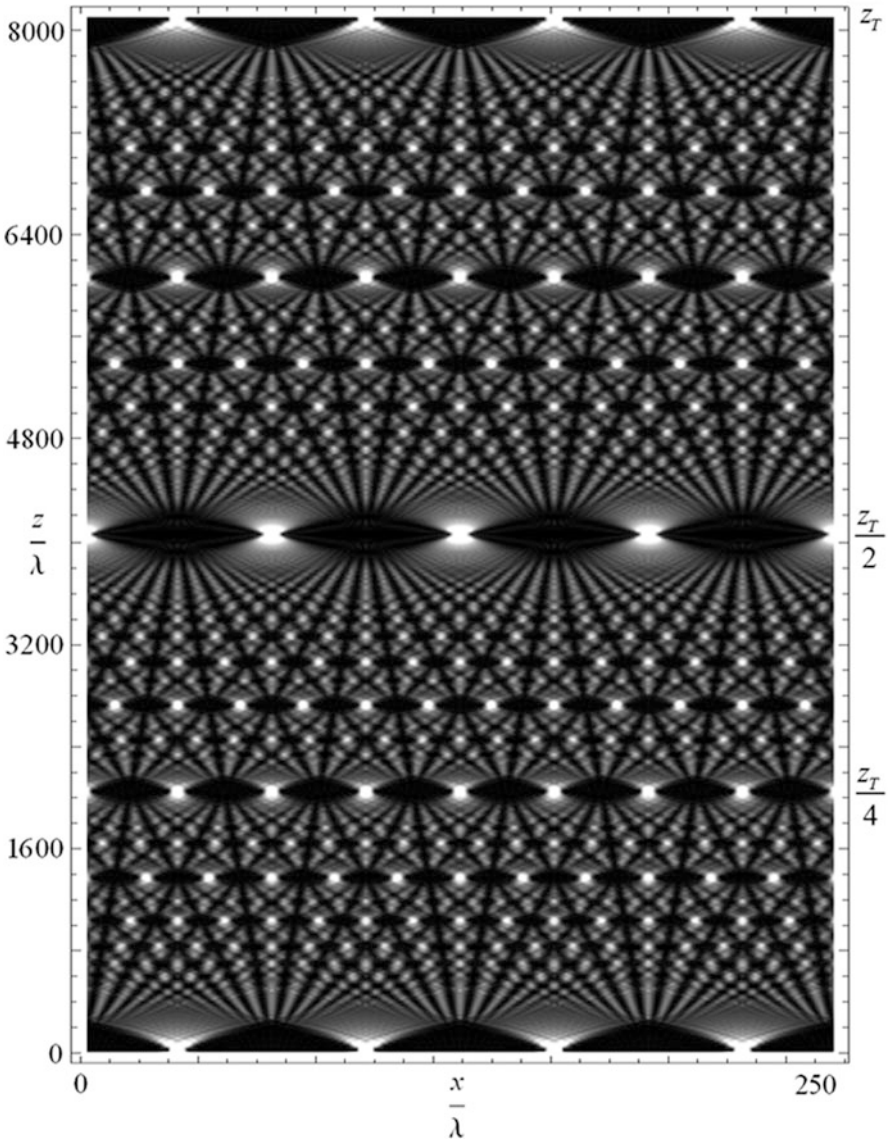


Fig. 2.46 Optical interference images of a grating mask due to Talbot effect

can be smaller than the dimension of mask opening, which was later named after him as the Talbot effect [112]. The Talbot effect is actually the interference of waves behind the grating mask, as illustrated in Fig. 2.46 where light waves coming through the mask openings interfere with each other and form black and white nodes. As seen in Fig. 2.46, at Z_T distance from the mask, the interference images

of same pitch as the mask grating reappear itself; at $Z_T/2$ position the interference images are also of the same pitch but have a 1/2 pitch shift laterally. The interesting phenomenon is that at the $Z_T/4$ position, the interference images have half the pitch of original mask grating or double the resolution. This is the key that Talbot effect can obtain high resolution grating patterns with a low resolution mask. These particular distances where the interference images are formed are called Talbot length and are expressed as Eq. 2.8, where a is the period of grating and λ is the wavelength of incident light.

$$Z_T = \frac{2a^2}{\lambda} \tag{2.8}$$

The Talbot effect can be utilized to pattern structures much smaller than mask features. The implementation of Talbot optical lithography is rather simple. It is similar to the proximity optical lithography using a mask aligner. The only difference is that the distance between mask and substrate has to be precisely controlled [113]. It was lately found that the distance between the mask and the substrate could be changed by a Talbot period during the exposure without altering the aerial image. So it does not require precise positioning at the Talbot distances but only a displacement of the mask or substrate by a Talbot period independently of the start and end positions. The new optical lithography technique is called displacement Talbot lithography (DTL), which has been commercialized based on the setup of mask aligner [114]. 75 nm gratings have been achieved by the technique with a 150 nm mask patterns exposed by a deep UV source of 266 nm wavelength [115]. Complex patterns, as long as they are periodical, can also be imaged by Talbot lithography [116]. Though it is simple, the Talbot lithography has limited capability for high resolution lithography. It only works for periodical patterns. The contrast of Talbot image is poor and for sub-100 nm patterning, high resolution mask and short wavelength illumination are necessary.

2.10 Interferometric Optical Lithography

Another simple, but capable of sub-100 nm resolution, and photon-based lithography is interferometric optical lithography. Optical interference refers to the interaction of light waves which are correlated or coherent with each other, either because they come from the same source or because they have the same or nearly the same wavelengths. The most famous example of optical interference is Young's double-slit experiment. Two beams from the double slit are both of same wavelength and same phase because they emit from the same light source. The interference of the two beams causes distribution of light intensity with periodical maxima and minima. Interferometric lithography (IL) is the use of a small number of coherent optical beams (minimum of two) incident from different directions onto a 2D (thin) or 3D (thick) photosensitive layer to produce an interference pattern whose intensity

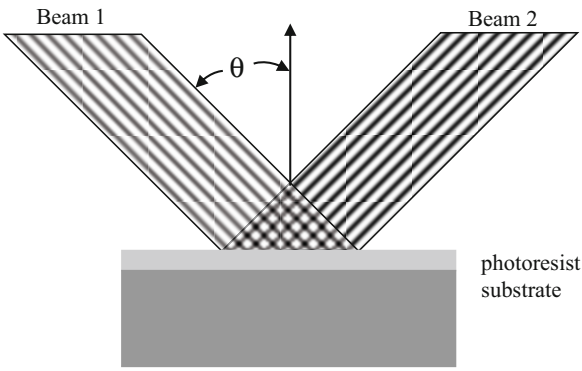


Fig. 2.47 Schematic of interferometric lithography

distribution is recorded in the photosensitive layer, as illustrated in Fig. 2.47. IL is a well-established lithographic technique. The use of IL for photoresist patterning goes back to the 1970s when it was mainly used to produce hologram plates [117, 118]. The alternative names are interference lithography or holographic lithography.

The intensity of interference pattern at the photoresist surface can be expressed as Eq. 2.9 [119].

$$I(x) = 2|E|^2\left[1 + \sigma_{\text{pol}}\cos(2kx\sin\theta)\right], \quad k = \frac{2n\pi}{\lambda} \tag{2.9}$$

where $\sigma_{\text{pol}} = 1$ for TE (transverse-electric) polarization or $\sigma_{\text{pol}} = \cos(2\theta)$ for TM (transverse-magnetic) polarization. λ is the illumination wavelength, θ is the angle of incident beam relative to normal as indicated in Fig. 2.47, and n is the refractive index. For light passing through air into photoresist, $n = 1$. The period of the interference pattern along the x direction is $\lambda/2n\sin(\theta)$. For a fixed wavelength, periodical lines with different pitches can be obtained by varying the angle of incident beams. At a fixed angle of incidence, further increase of resolution can be achieved by using short wavelength of illumination. It is apparent that at the maximum oblique angle of incidence ($\theta = 90°$), the minimum possible pitch is $\lambda/2$, which represents the resolution limit of IL. For half-pitch feature, the limit is $\lambda/4$. Increasing n or decreasing λ can both increase the resolution. The refractive index can be increased by immersion lithography. There has been demonstration of 29.9 nm half-pitch lines by IL with the combination of 193 nm wavelength and high index fluid immersion imaging [41]. IL with EUV wavelength has demonstrated 12.5 nm half-pitch features [120]. The IL certainly provides a very simple solution to push down further the resolution of optical lithography without using a mask and its associated complications such as phase-shifting and OPC.

The basic setup for IL is very simple. It only needs a laser, some lenses, and mirrors [121]. Figure 2.48 shows schematically a setup of IL system. A laser beam is

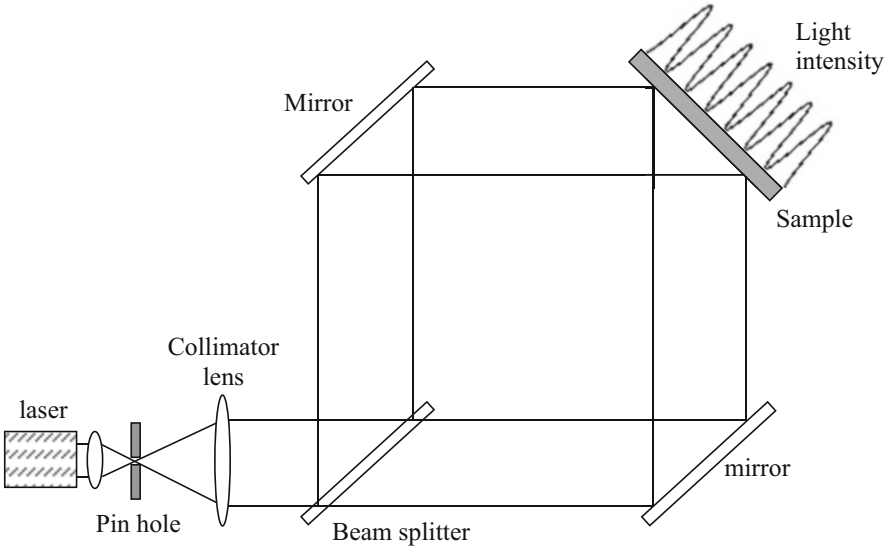


Fig. 2.48 Schematic of interferometric lithography setup

focused through a pinhole (to improve coherence) and collimated through a collimator lens and then split into two beams which are then reflected by two mirrors onto a photoresist-coated substrate. An interference light intensity distribution is formed and the photoresist is exposed by the interference light. For the lasers, early work was based on continuous-wave (cw) single-mode lasers such as the Ar ion laser at 488 nm (Ar+) and the He-Cd laser at 442 nm. With the commercial development of i-line lithography (365 nm), attention shifted to the cw Ar++ lines at 364 and 352 nm and to the 355 nm third harmonic of a YAG laser to take advantage of the availability of commercial i-line resists. Further pursuing of short wavelength has seen the use of the intra-cavity doubled Ar ion line at 244 nm and fifth-harmonic YAG at 213 nm. At wavelengths shorter than 257 nm, it is difficult to obtain sources with temporal and spatial coherence and output power sufficient for conventional IL. ArF excimer laser has the necessary output power but is poor in temporal and spatial coherence. Hence, to expose gratings and grids of sub-100 nm period, achromatic interferometric lithography (AIL) is used. AIL utilizes a group of phase grating plates to modulate the incoming laser light, so that the spatial period of the printed grating is dependent only on the period of the parent gratings used in the interferometer regardless of the optical path or the wavelength and coherence of the source [122].

IL has two distinctive advantages: One is its simplicity. In fact, any laboratories can install an IL system on an optical bench. Another is no need for a mask. The interference pattern is generated directly on a photoresist-coated substrate. The downside is that the interference pattern can only be in the form of straight lines from a single exposure, or dots if two exposures are carried out on the same substrate

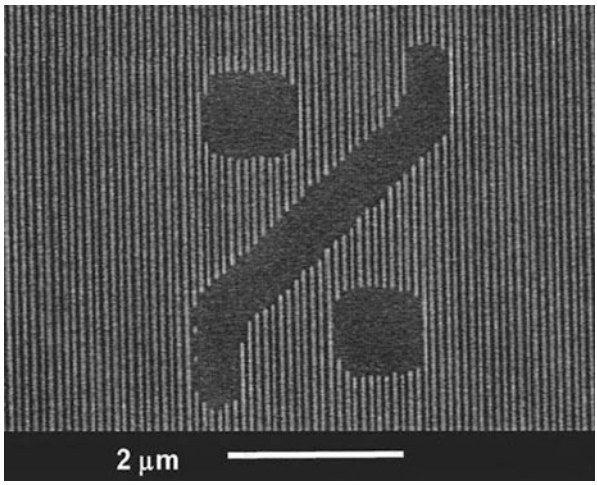


Fig. 2.49 Photoresist pattern produced by interference and projection hybrid lithography. (Reprint from Ref. [126] with permission)

with the second exposure rotating 90° to the first one. Whether straight lines or dots, they are periodical patterns with the pitch determined by the beam incident angle or wavelength of the laser. Although IL can only pattern periodical features, there are many applications which only require periodical features, for example, 2D and 3D photonic crystal structures [123], magnetic storage devices [124], etc. There was a report using seven laser beams shining from all directions through a thick lump of SU-8 photoresist, and complex 3D interference structures were formed as a photonic bandgap material [125].

IL is not completely useless in IC manufacturing. It can be adapted to patterning IC design layouts. One of the ideas is hybrid maskless optical lithography, which uses IL to produce periodical patterns first and then uses a conventional photomask to trim the interference patterns into desired arbitrary pattern shapes. Figure 2.49 gives an example of the hybrid lithography, where a line/space pattern has been trimmed in the middle with a coarse dark field mask. The trim mask is usually of low resolution. One can work out many ways of designing a trim mask to tailor the interference pattern into a desired IC layout. The key advantage of the hybrid approach is that the most difficult part of lithography, i.e., the highest resolution patterning, is done by the maskless IL, which saves many troubles from using those RET masks mentioned earlier.

The IL should, however, not be confused with imaging interferometric lithography (IIL) technique. IIL is basically a projection optical lithography with a photomask but combined with interference from a reference beam to enhance the resolution of projection lithography [127].

2.11 Maskless Optical Lithography

"The devil is in the mask." This was the message delivered by Burn Lin, the world renowned optical lithography expert, at the SPIE 20th Annual Conference on Optical Lithography in March 2007 [128]. That message still holds true at the present day. Optical lithography, since it started from contact printing, has always been relying on photomasks to deliver pattern designs to substrates. With IC manufacturing going into sub-50 nm feature dimension, photomasks have increasingly become a bottleneck as more and more RET features are incorporated into the photomasks, such as phase-shifting, SRAF, OPC, and masks generated by inverse lithography, as have been described in previous sections. With the introduction of multi-patterning, the number of masks needed has also increased considerably. For example, at 16 nm node the total number of masks is as high as 75 [129]. The adoption of EUV lithography has reduced the number of masks, but total cost of a mask set has greatly increased. Ten years ago, at the most advanced nodes, a leading-edge mask set is at the price of $2 million to $3 million, or roughly $100,000 for a single mask at critical feature level [130]. Today at 7 nm node, a single EUV mask can cost over $0.5 million and the whole set of masks costs up to $15 million. The escalating mask cost can have serious impact on the overall cost of low-volume IC production such as foundry runs and fabrication of application-specific integrated circuits (ASICs). For small-volume IC manufacturing, maskless lithography (ML2) was favored [131]. ML2 includes a number of optical and nonoptical technologies. The nonoptical ML2 techniques use electron or ion beams, which will be described in Chaps. 3 and 4. For optical ML2, a number of techniques exist. The interferometric lithography (IL) introduced in the previous section is a ML2 technique. However, as mentioned above, IL does not offer the total flexibility because only periodical patterns can be generated by IL. In this section, some other optical ML2 techniques are introduced.

Maskless optical lithography has been around for more than three decades. It was started in 1987 by the introduction of first laser direct write tool [132]. It used a scanning laser beam to perform direct exposure of photoresist. The technology deemed to be slow and low throughput, therefore, was not able to compete with conventional projection optical lithography. However, it was highly suitable for lithography in making photomasks. At the time, mask making was dominantly using electron beam (e-beam) raster scan system. The first laser tool CORE 2000 offered multi-laser beams with raster scan, which greatly increased the throughput of exposure. Since then, laser direct write tools have evolved from the first generation of CORE 2000, CORE 2564 to ALTA 3000, ALTA 4300, and the latest model of ALTA 4700DP supplied by Applied Materials Inc. While the ALTA 4700 has comparable resolution as a 50 keV shaped e-beam mask writer, the writing time is nearly half of that of a 50 keV e-beam writer [133]. Over 75% of masks are now produced by laser tools.

A simple schematic drawing of direct laser scan is shown in Fig. 2.50. A single laser beam from a frequency-doubled continuous-wave (cw) Ar ion laser at

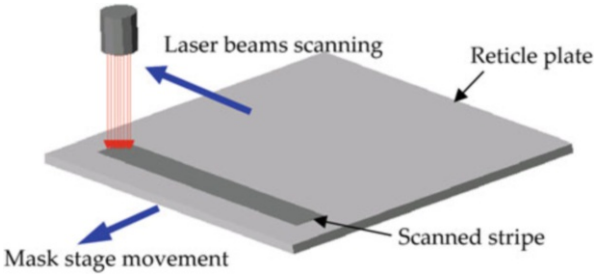


Fig. 2.50 Schematic of laser scan in making reticles

wavelength of 257 nm is split into 32 individually addressable beams, like a light brush. The brush is scanned across the plate by a 24-facet polygonal mirror. A single facet of polygon mirror scans the brush over an area of 4096 pixels wide by 32 pixels long. While the brush is scanned across the reticle by the facet of mirrors, the reticle stage is driven by a linear motor stage moving at constant velocity in the direction perpendicular to the scan. A set of brush strokes made during one stage move is called a stripe. A stripe is always 4096 pixels wide, but its length varies up to the length of the reticle depending on the pattern data. After a stripe is written, the stage performs a rapid retrace and moves to the starting position of the next stripe. This stage movement lines up each successive "brush stroke" to provide complete area coverage. One should remember that a mask feature is 4 times larger than the feature on a wafer, because of the 4:1 reduction of project optical lithography system. For ALTA 4700, the pixel size is 96 nm in the x-axis and 160 nm in the y-axis. Therefore, the resolution capability of a laser direct tool is adequate for patterning the masks for latest generation of IC manufacturing, including all the sub-resolution features used in RET.

In parallel to the aforementioned multi-beam laser tools, there is another class of optical maskless lithography which is based on micromirror array to reflect or deflect light [134]. This new scanning scheme was benefited from the development of MEMS. A micromirror is a MEMS device in which the mirror surface is tilted by applying electrostatic field between the mirror plate and an underlying electrode, as schematically shown in Fig. 2.51a. The dimension of a mirror is only a few tenths of square micrometers. A large closely packed mirror array can be fabricated by silicon surface micromachining process, such as the digital micromirror device (DMD) mass produced by Texas Instruments (TI). A typical DMD from TI (DLP4500) is measured 11.43 mm diagonal with 912 × 1140 array of aluminum-coated micromirrors and 7.6 μm mirror pitch. The micromirrors can have ±12° tilt angle (relative to flat state). Because the deflection of each micromirror is individually controlled by a pair of underlying electrodes, it acts as a pixel. Light shining onto the mirror surface can be reflected to different angles, depending on the deflection of mirror surface. An aperture in an optical imaging system can control which angle of reflected light can pass through. A flat mirror represents a "white" pixel, and a mirror deflected by one fourth of wave length at the edge of mirror represents a "black" pixel. Deflection in between represents a "gray" pixel, as schematically illustrated in

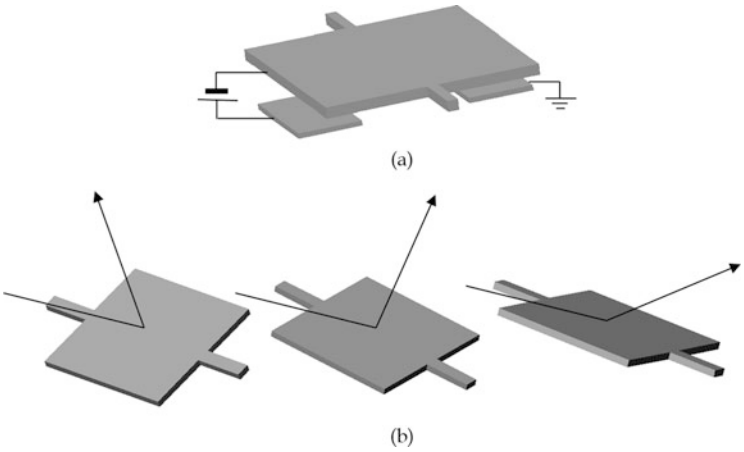


Fig. 2.51 (**a**) Micromirror structure, (**b**) white, gray, and black pixels, represented by different deflections of mirror surface

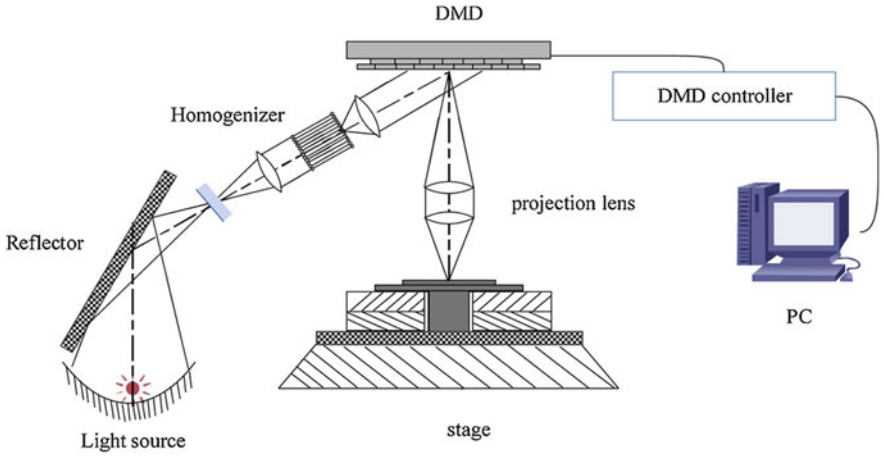


Fig. 2.52 Setup of micromirror-based maskless lithography system. (Reprinted from [137] with permission)

Fig. 2.51b. Therefore, the micromirror array is commonly called spatial light modulator (SLM). There are other types of SLM that have been adopted for optical maskless lithography, including grating light valve (GLV) for phase modulation [135] and liquid crystal light valve for intensity modulation [136].

The micromirror-based laser lithography tool works like an optical stepper. Figure 2.52 shows a schematic of exposure system setup [137]. A DMD chip, consisting of an array of micromirrors, is controlled by a pattern generator in which pattern data is converted into electrical signal to control the deflection of

each micromirror. The SLM is uniformly illuminated by light from a pulsed excimer laser which has been processed through a beam homogenizer. The reflected light pattern is imaged through a demagnifying optical system onto the photoresist-coated substrate. An interferometer-controlled air-bearing stage moves continuously in the x direction, and the laser is triggered at the exact time to stitch the exposed field with the previous exposure. In the time between two sequential exposures, the SLM is reloaded with new data corresponding to the pattern to be written. After a stripe of fields has been exposed over the full length of the substrate, the stage moves in the y direction, and a new stripe is exposed next to the previous one.

The first generation of SLM laser tool had reduction of 160× [134]. Later it was developed to reduction of 267×. With a micromirror size of 8 × 8 μm, the achievable minimum pixel size was 30 nm [138]. The commercial SLM-based lithography tools are represented by the Sigma-7000 series from Micronic Laser System AB. Their Sigma-7500 tool was able to make masks for 45 nm generation of IC manufacturing [139]. However, mask making for IC industry is a very competitive market. With only a few companies left for manufacturing the latest generations of ICs, only the ALTA tools from Applied Materials Inc. survived to serve the IC industry, while other vendors such as the Micronic and Heidelberg Instruments are providing maskless laser lithography tools for a broad scope of applications other than ICs, for example, masks for making display panels [140] or lithography for microelectromechanical systems (MEMS) using the so-called maskless aligner as opposed to the traditional mask aligner [141]. Other microscale patterning needs include microoptics, microfluidics, computer-generated hologram (CGH), etc. The pattern features are no longer in nanoscale but in micrometer size, with emphasis on large substrates which can be up to 1400 mm × 1620 mm for display applications [142].

2.12 Two-Photon Polymerization Lithography

While the above optical ML2 technologies are well developed by industry and very much geared toward volume production of photomasks or ICs, there is another laser-based lithography technique developed in research laboratories, which can directly form 3D structures without using a photomask. This is the two-photon polymerization (2PP) lithography. 2PP is based on photon-induced polymerization process to solidify a liquid polymer upon being irradiated by a laser beam. It is very similar to a microscale photon polymerization technique, called microstereolithography, which is commonly referred to as 3D printing nowadays [143]. While microstereolithography uses a UV laser beam to scan a surface layer of liquid polymer to form a 2D pattern and build up 3D structures layer by layer, 2PP lithography uses infrared femtosecond laser beams to form a focus point inside a pool of liquid polymer, and the polymerization only takes place at the focal point.

Microstereolithography is a single-photon polymerization process, in which the short wavelength of UV laser has enough energy to trigger the polymerization of

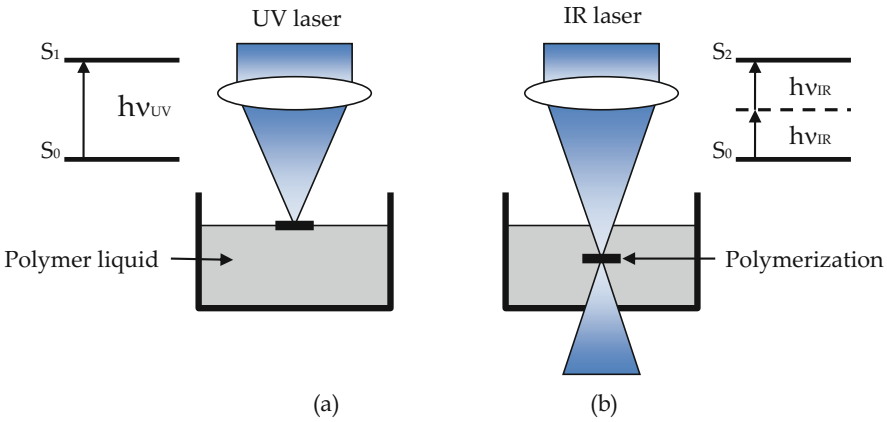


Fig. 2.53 Schematic comparison of (**a**) microstereolithography (one-photon polymerization) and (**b**) two-photon polymerization lithography

some photon-sensitive polymers. Therefore, the UV laser polymerization only takes place at the surface of liquid polymer, as illustrated in Fig. 2.53a. A 3D object is built up by lowering the polymer liquid surface layer by layer. In two-photon polymerization process, the energy of a single infrared photon is not enough to trigger the polymerization. Two photons that overlap double the energy or shorten the wavelength by half, which is equivalent to the absorption of UV photons as illustrated in Fig. 2.53b. The polymerization is only triggered at the location of overlapped two photons [144]. The two-photon polymerization is only possible with a femtosecond infrared laser at its focal point. The polymerization can happen inside the liquid polymer because the polymer is transparent to infrared laser.

By moving around the two-photon focus point, any 3D polymerized structures can be formed. The nonlinear absorption of photons with controllable threshold intensity allows the 2PP lithography to have very high resolution at the level of $\lambda/10$ to $\lambda/50$ [145]. Making 3D structures with sub-100 nm details is possible. Figure 2.54 shows a chain-like 3D structure formed by two-photon polymerization of a liquid resin. The smallest structure dimension is only 120 nm. 2PP lithography can induce not only polymerization but also variety of other photochemical processes. For example, 3D metallic microstructures were successfully created in the polymer matrix containing homogeneously doped metal ion by photon-induced reduction process [146], as well as a variety of other materials [147]. Commercial 2PP lithography tools, the Quantum X series, are available at the German company Nanoscribe [148].

The key condition to realize 2PP is that the laser beams do not cause significant absorption of photon energy in the liquid polymer except at the focus point of two laser beams. In the past infrared (IR) lasers were used for 2PP lithography because only IR wavelength has the minimum absorption in the polymer. Recently, new photon-sensitive polymers have been discovered, which can have peak absorption of

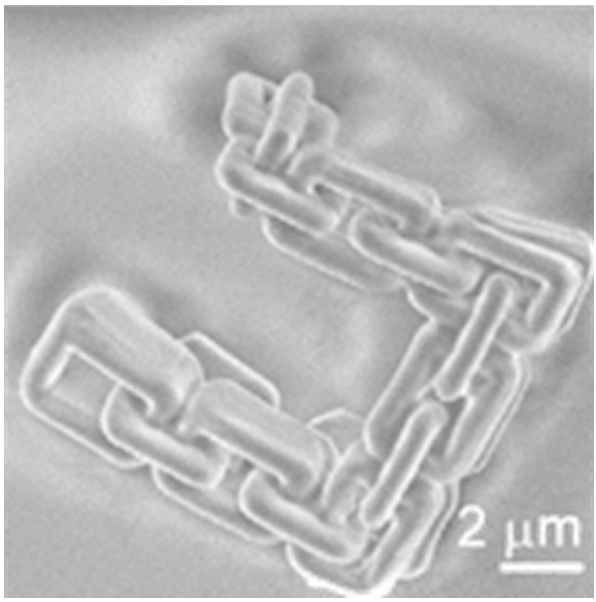


Fig. 2.54 A chain-like 3D structure formed by two-photon polymerization of liquid resin. (Reprint from Ref. [149] with permission)

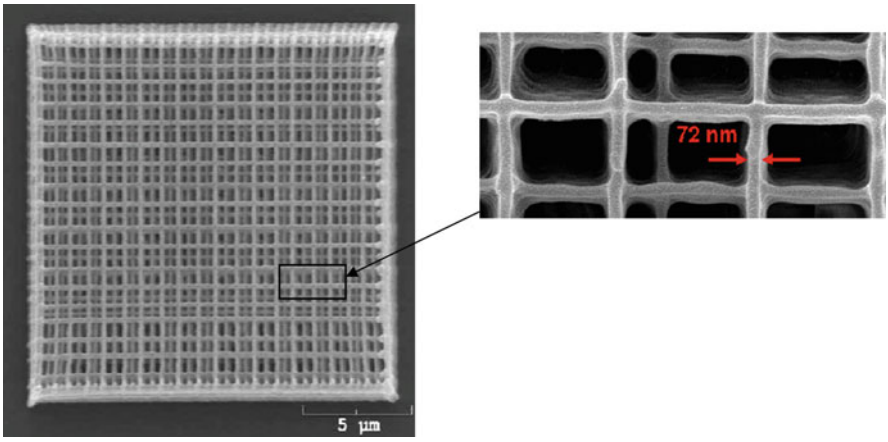


Fig. 2.55 "Woodpile" structure formed by two-photon exposure of liquid polymer. (Reprint from Ref. [150] with permission)

photon energy at shorter wavelength. Visible femtosecond lasers at wavelengths of 520 nm and 730 nm were used. By controlling the laser exposure dose, 3D polymer "woodpile" structures have been formed with the minimum line diameters around 60–70 nm, as is shown in Fig. 2.55. Such structure can be useful as a photonic crystal material.

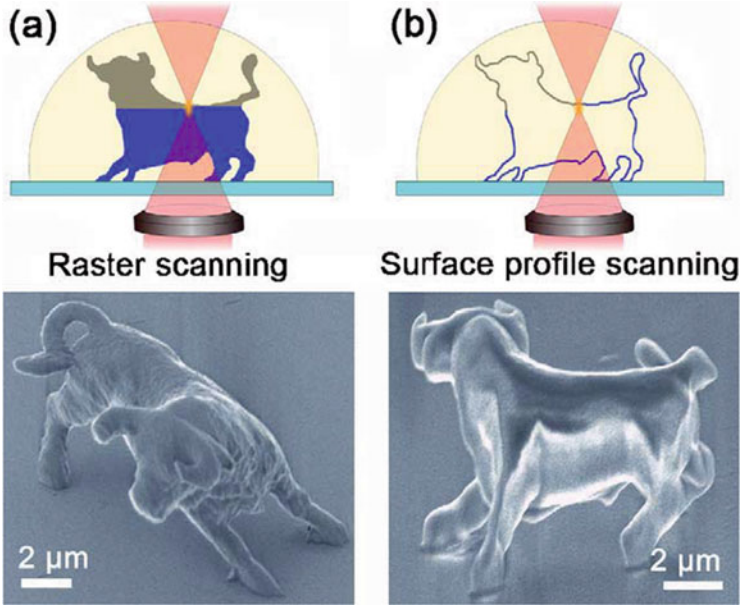


Fig. 2.56 Polymerization of a bull sculpture by (**a**) raster scan of the whole body or (**b**) only the surface shell and subsequent solidification by flood UV exposure from a mercury lamp. (Reprint from Ref. [151] with permission)

Though complicated 3D polymer structures can be made by 2PP lithography, the process is extremely slow and inefficient because the structure is formed point by point with focused laser. One remedy for this problem is to form a shell of 3D object and then polymerize the inside by flood exposure of UV light, as the example shown in Fig. 2.56. For the bull sculpture to completely solidify by 2PP raster scan, it would take 183 min. By polymerization of only the surface profile (the shell) and subsequently solidifying by a mercury lamp, the same bull sculpture could be made in 13 min [151].

References

1. Lin, B.: The ending of optical lithography and the prospects of its successors. Microelectron. Eng. **83**, 604–613 (2006)
2. Cui, Z.: Nanofabrication, Principles, Capabilities and Limits, 2nd edn. Springer (2016)
3. Cui, Z.: Micro-Nanofabrication Technologies and Applications. Springer (2006)
4. Wong, A.K.: Optical Imaging in Projection Microlithography. SPIE Press (2005)
5. Cui, Z.: Computer simulation of optical lithography in VLSI fabrication. In: Symposium of Chinese Young Scholars on Optical Application Technologies (1994)
6. De Simone, D., et al.: Potential applications of negative tone development in advanced lithography. Microelectron. Eng. **88**, 1917–1922 (2011)

7. Ito, H.: Deep-UV Resists: Evolution and Status, p. 164. Solid State Technology (1996)
8. International Technology Roadmap for Semiconductors (ITRS). Available from: http://www.itrs.net/
9. Bakshi, V.: EUV Sources for Lithography. SPIE Press (2006)
10. Yang, D.-K.: The development of laser-produced plasma EUV light source. Chip. **1**, 100019 (2022)
11. Purvis, M., et al.: Industrialization of a robust EUV source for high-volume manufacturing and power scaling beyond 250W. In: Advanced Lithography: Extreme Ultraviolet (EUV) Lithography IX. SPIE (2018)
12. Brandt, D.C., et al.: Performance and availability of EUV sources in high volume manufacturing on multiple nodes in the field and advances in source power. SPIE Proc. **11854** (2021)
13. Umstadter, K., et al.: EUV light source for high-NA and low-NA lithography. SPIE Proc. **12494** (2023)
14. Tomofuji, T., et al.: Mo/Si multilayer (ML) mirror deposited with ion beam sputtering using Kr gas. In: 3rd International EUVL Symposium (2004)
15. Stoeldraijer, J., Graeupner, P., Kuerz, P., Stammler, T., van Schoot, J.: EUV optics: status, outlook and future. SPIE Opt. EUV Nanolithogr. XXXV. **12051** (2022)
16. Kim, J.S., Jinho, A.: Mask materials and designs for extreme ultra violet lithography. Electron. Mater. Lett. **14**, 533–547 (2018)
17. Chris Progler, M.G., Bonam, R.: EUV mask challenges and requirements for ultimate single exposure interconnects. SPIE Proc. **10957** (2019)
18. Stivers, A., et al.: EUV mask defects strategy. In: EUV Mask Development Seminar (2002)
19. Naulleau, P., et al.: Lithographic characterization of the printability of programmed extreme ultraviolet substrate defects. J. Vac. Sci. Technol. **B21**(4), 1286 (2003)
20. Tilmann Heil, M.W., Capelli, R.: Pushing the limits of EUV mask repair: addressing sub-10 nm defects with the next generation e-beam-based mask repair tool. J. Micro/Nanopatterning Mater. Metrol. **20**(3), 031013 (2021)
21. Theodore Manouras, P.A.: High sensitivity resists for EUV lithography: a review of material design strategies and performance results. Nanomaterials (Basel). **10**, 1593 (2020)
22. Chaoyun Luo, X.H., Liu, W.: Review of recent advances in inorganic photoresists. RSC Adv. **10**, 8385–8395 (2020)
23. Goldfarb, D.L., Lewis, S.M., Grindell, R., Timco, G.A., Winpenny, R.E.P.: Dry heterometallic resist processing based on thermal sublimation deposition and development. J. Micro/Nanopatterning Mater. Metrol. **21**(4), 041406 (2022)
24. Yen, A., Miyazaki, J.: EUV lithography technology for high-volume production of semiconductor devices. J. Photopolym. Sci. Technol. **32**(2), 195–201 (2019)
25. Wind, S., et al.: Lithography and fabrication processes for sub-100 nm scale complementary metal–oxide semiconductor. J. Vac. Sci. Technol. **B13**(6), 2688 (1995)
26. Krasnoperva, A.A., et al.: Imaging capability of proximity X-ray lithography at 70nm ground rule. Proc. SPIE. **3676**, 24 (1999)
27. Maldonado, J.R., Peckerar, M.: X-ray lithography: some history, current status and future prospects. Microelectron. Eng. **161**, 87–93 (2016)
28. Abrar Faisal, T.B., Mohr, J.: Influence of secondary effects in the fabrication of submicron resist structures using deep x-ray lithography. J. Micro/Nanolithogr. MEMS MOEMS. **18**(2), 023502 (2019)
29. Becker, E.W., et al.: Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming and plastic moudling (LIGA process). Microelectron. Eng. **4**, 35 (1986)
30. Lin, B.J.: The k3 coefficient in nonparaxial lambda/NA scaling equations for resolution, depth of focus, and immersion lithography. J. Microlith. Microfab. Microsys. **1**, 7 (2002)
31. Wim de Boeij, G.D., de Wit, M.: Extending KrF lithography beyond 80nm with the TWINSCAN XT:1000H 0.93NA scanner. SPIE Proc. **7140** (2008)

32. Fay, B.: Advanced optical lithography development, from UV to EUV. Microelectron. Eng. **61–62**, 11–24 (2002)
33. Technology backgrounder: Immersion Lithography in IC Knowledge. 2003
34. Smith, B.W., Cashmore, J.: Challenges in high NA, polarization, and photoresists. Proc. SPIE. **4691** (2002)
35. McCallum, M., Fuller, G., Owa, S.: From hyper NA to low NA. Microelectron. Eng. **83**, 667–671 (2006)
36. Soichi Owa, H.N.: Immersion lithography: its history, current status and future prospects. SPIE Proc. **7140** (2008)
37. Chang, C.-Y., et al.: Watermark defect formation and removal for immersion lithography. SPIE Proc. **6154** (2006)
38. Peters, L.: Photoresists Meet the 193 nm Milestone. Semiconductor International (2005)
39. Wei, Y., Petrillo, K., Benson, P.A.: Evaluating Topcoat Options for Immersion Litho Resists. Semiconductor International (2006)
40. Wu, S., et al.: Non-topcoat resist design for immersion process at 32-nm node. Proc. SPIE. **6923** (2008)
41. IBM: JSR: Immersion Can Extend Past 32nm Node. Solid State Technology (2006)
42. Zimmerman, P.A., et al.: High index 193 nm immersion lithography: the beginning or the end of the road. Proc. SPIE. **7274** (2009)
43. Abercrombie, D.: Will EUV Kill Multi-Patterning? In: Semiconductor Engineering (2017)
44. Yen, A.: Rayleigh or Abbe? Origin and naming of the resolution formula of microlithography. J. Micro/Nanolithogr. MEMS MOEMS. **19**(4), 040501 (2020)
45. Tounai, K., et al.: Resolution improvement with annular illumination. Proc. SPIE. **1674**, 1753 (1992)
46. Shiraishi, N., et al.: New image technique for 64M-DRAM. Proc. SPIE. **1674**, 1741 (1992)
47. Herkommer, A.M.: Evolution of illumination Systems in Microlithography: a retrospective. Proc. SPIE. **7652** (2010)
48. Melchior Mulder, A.E., Noordman, O.: Performance of a programmable illuminator for generation of freeform sources on high NA immersion systems. SPIE Proc. **7520** (2009)
49. Thuc Dam, V.T., Hu, P.: Source-mask optimization (SMO): from theory to practice. SPIE Proc. **7640** (2010)
50. Levenson, M.D., Viswanathan, N.S., Simpson, R.A.: Improving resolution in photolithography with a phase-shifting mask. IEEE Trans. Electron. Devices. **ED-29**, 1828 (1982)
51. Pelka, J., Henke, W.: Simulation and optimization of phase-shift masks for printing of contact holes. Microelectron. Eng. **26**, 1–26 (1994)
52. Cui, Z., Prewett, P.D., Johnson, S.: Transmission and side-lobe effect in attenuated phase shift masks. Microelectron. Eng. **27**, 259 (1995)
53. Hand, A.: Advanced Masks Help Keep Photolithography Alive. Semiconductor International (2007)
54. Andreas Erdmann, H.M., Evanschitzky, P.: Attenuated phase shift masks: a wild card resolution enhancement for extreme ultraviolet lithography? J. Micro/Nanopatterning Mater. Metrol. **21**(2), 020901 (2022)
55. Van Den Broeke, D., et al.: Complex 2D pattern lithography at λ/4 resolution using chromeless phase lithography (CPL). Proc. SPIE. **4691**, 196 (2002)
56. Chen, J.F., et al.: Manufacturing at k1 = 0.2 with Chromeless Phase Lithography. Solid State Technology (2002)
57. Levenson, M.D.: Extending the lifetime of optical lithography by wave-front engineering. Jpn. J. Appl. Phys. **33**(12B), 6765 (1994)
58. Cui, Z., et al.: Optical proximity correction by grey tone photolithography. Microelectron. Eng. **53**, 153 (2000)
59. Mack, C.A.: Scattering Bars. Solid State Technology (2003)
60. Otto, O.W., et al.: Automatic optical proximity correction: a rule-based approach. Proc. SPIE. **2197**, 278 (1994)

61. Rieger, M.L., Stirniman, J.P.: Using behavior modelling for proximity correction. Proc. SPIE. **2197**, 371 (1994)
62. Wu, Q., Hou, Y.: Optical proximity correction, methodology and limitations. In: China Semiconductor Technology International Conference (CSTIC), vol. 1-5. IEEE, Shanghai (2021)
63. Cui, Z., Du, J.: Investigation of OPC mask distortion effect. Proc. SPIE. **4404** (2001)
64. Hector, S.: Behind the photomask. In: Global Semiconductor Forum (2005)
65. Saleh, B.E.A., Sayegh, S.I.: Reductions of errors of microphotographic reproductions by optical corrections of original masks. Opt. Eng. **20**, 781–784 (1981)
66. Pang, L.: Inverse lithography technology: 30 years from concept to practical, full-chip reality. J. Micro/Nanopatterning Mater. Metrol. **20**(3), 030901 (2021)
67. Pang, L., et al.: Study of mask and wafer co-design that utilizes a new extreme SIMD approach to computing in memory manufacturing: full-chip curvilinear ILT in a day. Proc. SPIE. **11148** (2019)
68. Pang, L., et al.: TrueMask ILT MWCO: full-chip curvilinear ILT in a day and full mask multi-beam and VSB writing in 12 hrs for 193i. Proc. SPIE. **113270** (2020)
69. Xu Ma, G.R.A.: Computational Lithography. Wiley (2010)
70. eBeam Initiative 9th annual Survey Results. https://www.ebeam.org/docs/2022-ebeam-luminaries-survey-final-1.pdf
71. Shin, Y.: Computational lithography using machine learning models. IPSJ Trans. Syst. LSI Des. Methodol. **14**, 2–10 (2021)
72. Pang, L., Liu, Y., Abrams, D.: Inverse lithography technology (ILT): what is the impact to the photomask industry? Proc. SPIE. **6283** (2006)
73. Rai-choudhury, P. (ed.): Handbook of Microlithography, Micromachining and Microfabrication, vol. 1. SPIE Press (1997)
74. Zant, P.V.: Microchip Fabrication, 5th edn. McGraw-Hill (2004)
75. Ito, H., Willson, C.G.: In: Davidson, T. (ed.) Polymers in Electronics Symposium Series 242, p. 11. American Chemical Society, Washington, DC (1984)
76. Reichmanis, E., Thompson, L.F.: Chemistry and processes for deep-UV resists. Microelectron. Eng. **13**, 3 (1991)
77. Ruede, D., Ercken, M., Borgers, T.: The Impact of Airborne Molecular Bases on DUV Photoresists. Solid State Technology (2001)
78. Slezak, M.: Multilayer Resist Strategies. Solid State Technology (2003)
79. Mack, C.A.: The Lithography Expert: Line-Edge Roughness, Part 1. Microlithography World (2007)
80. Cutler, C.A., Mackevich, J.F., Li, J.: Effect of polymer molecular weight on AFM polymer aggregate size and LER of EUV resists. Proc. SPIE. **5037**, 406 (2003)
81. Gallatin, G.M.: Resist blur and line edge roughness. Proc. SPIE. **5754**, 38 (2005)
82. Jiang, X., et al.: Investigations on the correlation between line-edge-roughness (LER) and line-width-roughness (LWR) in nanoscale CMOS technology. In: 11th International Conference on Solid-State and Integrated Circuit Technology. IEEE, Xi'an (2012)
83. Wang, R., et al.: Investigations on line-edge roughness (LER) and line-width roughness (LWR) in nanoscale CMOS technology: part II–experimental results and impacts on device variability. IEEE Trans. Electron Devices. **60**(11), 3676–3682 (2013)
84. Thakur, N., et al.: Stability studies on a sensitive EUV photoresist based on zinc metal oxoclusters. J. Micro/Nanolithogr. MEMS MOEMS. **18**(4), 043504 (2019)
85. Luo, C., et al.: Review of recent advances in inorganic photoresists. RSC Adv. **10**, 8385 (2020)
86. Kim, B.-G., et al.: Beneath the MEEF. Solid State Technology (2000)
87. Su, B., et al.: Analyzing and Characterizing 193nm Resist Shrinkage. Solid State Technology (2001)
88. Czech, G., Richter, E., Wunnicke, O.: 193nm resists: a status report (part one). Future Fab Intl. **12** (2002)

89. Tanaka, T., Morigami, M., Atoda, N.: Mechanism of resist pattern collapse during development process. Jpn. J. Appl. Phys. **32**, 6059 (1993)
90. Cao, H.B., Nealey, P.F.: Comparison of resist collapse properties for deep ultraviolet and 193 nm resist platforms. J. Vac. Sci. Technol. **B18**, 3303 (2000)
91. Ye, J., et al.: Prevention of photoresist pattern collapse by using liquid carbon dioxide. Ind. Eng. Chem. Res. **40**(24), 5858–5860 (2001)
92. Owa, S., et al.: Immersion lithography extension to sub-10 nm nodes with multiple patterning. Proc. SPIE. **9052**, 905200-1 (2014)
93. Miller, M.: Yield-Aware Designs Target Production Issues. Solid State Technology (2005)
94. Kobayashi, S., et al.: Automated hot-spot fixing system applied to the metal layers of 65-nm logic devices. J. Micro/Nanolithogr. MEMS MOEMS. **6**, 031010 (2007)
95. Ho, J., et al.: Lithography-simulation-based design for manufacturability rule development: an integrated circuit design house's approach. J. Micro/Nanolithogr. MEMS MOEMS. **6**, 031008 (2007)
96. Mansfield, S., Han, G., Liebmann, L.: Through-process modeling for design-for-manufacturability applications. J. Micro/Nanolithogr. MEMS MOEMS. **6**, 031007 (2007)
97. Yeric, G., et al.: The past, present, and future of design-technology co-optimization. In: The IEEE 2013 Custom Integrated Circuits Conference. IEEE, San Jose (2013)
98. Zhou, W., et al.: DFM: "Design for Manufacturing" or "design friendly manufacturing". J. Microelectron. Manuf. **3**, 20030101 (2020)
99. Blaikie, R.J., et al.: Nanolithography using optical contact exposure in the evanescent near field. Microelectron. Eng. **46**, 85 (1999)
100. Alkaisi, M.M., et al.: Sub-diffraction-limited patterning using evanescent near-field optical lithography. Appl. Phys. Lett. **75**(22), 3560 (1999)
101. Blaikie, R., Hong, F.: Plasmonic lithography: recent progress. Adv. Opt. Mater. **7**, 1801653 (2019)
102. Barnes, W.L.: Surface plasmon–polariton length scales: a route to sub-wavelength optics. J. Opt. A Pure Appl. Opt. **8**, S87–S93 (2006)
103. Luo, X., Ishihara, T.: Surface plasmon resonant interference nanolithography technique. Appl. Phys. Lett. **84**(23), 4780 (2004)
104. Goodberlet, J.G., Kavak, H.: Patterning Sub-50 nm features with near-field embedded-amplitude masks. Appl. Phys. Lett. **81**(7), 1315 (2002)
105. Fang, N., et al.: Sub–diffraction-limited optical imaging with a silver superlens. Science. **308**, 534 (2005)
106. Fang, L., et al.: Superlens imaging lithography for high aspect ratio sub-wavelength pattern employing trilayer resist process. Microelectron. Eng. **110**, 35–39 (2013)
107. Changtao Wang, X.L., et al.: Plasmonic structures, materials and lenses for optical lithography beyond the diffraction limit: a review. Micromachines. **7**, 118 (2016)
108. Zhao, Z., et al.: Going far beyond the near-field diffraction limit via plasmonic cavity lens with high spatial frequency spectrum off-axis illumination. Sci. Rep. **5**, 15320 (2015)
109. Wen, X., et al.: High throughput optical lithography by scanning a massive array of bowtie aperture antennas at near-field. Sci. Rep. **5**, 16192 (2015)
110. Paik, S., et al.: Near-field sub-diffraction photolithography with an elastomeric photomask. Nat. Commun. **11**, 805 (2020)
111. Kobrin, B., et al.: Rolling mask nanolithography: the pathway to large area and low cost nanofabrication. Proc. SPIE. **8249** (2012)
112. Talbot, W.H.F.: Facts relating to optical science. Philos. Mag. **9**, 403–405 (1836)
113. Dunbar, L.A., et al.: Talbot lithography as an alternative for contact lithography for submicron features. Proc. SPIE. **8974** (2014)
114. Solak, H.H., Dais, C., Clube, F.: Displacement Talbot lithography: a new method for high-resolution patterning of large areas. Opt. Express. **19**(11), 10686 (2011)
115. Wang, L., et al.: Sub-wavelength printing in the deep ultra-violet region using displacement Talbot lithography. Microelectron. Eng. **161**, 104–108 (2016)

116. Isoyan, A., et al.: Talbot lithography: self-imaging of complex structures. J. Vac. Sci. Technol. **B 27**, 2931 (2009)
117. Beesley, M.J., Castledine, J.G.: The use of photoresist as a holographic recording medium. Appl. Opt. **9**, 2720 (1970)
118. Zaidi, S.H., Brueck, S.R.J.: High aspect-ratio holographic photoresist gratings. Appl. Opt. **27**, 2999–3002 (1988)
119. Brueck, S.R.J.: Optical and interferometric lithography—nanotechnology enablers. Proc. IEEE. **93**(10), 1074 (2005)
120. Solak, H.H., et al.: Photon-beam lithography reaches 12.5 nm half-pitch resolution. J. Vac. Sci. Technol. **B25**, 91 (2007)
121. van Wolferen, H., Abelmann, L.: In: Hennessy, T.C. (ed.) Laser Interference Lithography in Lithography: Principles, Processes and Materials. Nova Science Publishers (2011)
122. Savas, T.A., et al.: Achromatic interferometric lithography for 100-nm-period gratings and grids. J. Vac. Sci. Technol. **B13**(6), 2732 (1995)
123. Campbell, M., et al.: Fabrication of photonic crystals for the visible spectrum by holographic lithography. Nature. **404**, 53–56 (2000)
124. Ross, C.A., et al.: Magnetic behavior of lithographically patterned particle arrays. J. Appl. Phys. **91**, 6848–6853 (2002)
125. Pang, Y.K., et al.: Chiral microstructures (spirals) fabrication by holographic lithography. Opt. Express. **13**(19), 7615 (2005)
126. Fritze, M., et al.: Hybrid optical maskless lithography: scaling beyond the 45 nm node. J. Vac. Sci. Technol. **23**(6), 2743 (2005)
127. Chen, X., Brueck, S.R.J.: Imaging interferometric lithography: a wavelength division multiplex approach to extending optical lithography. J. Vac. Sci. Technol. **B16**(6), 3392 (1998)
128. Resor, G.: Burn Lin's Wish List for Optical Lithography: Eliminate the Mask. Solid State Technology (2007)
129. LaPedus, M.: Photomask shortages grow at mature nodes. In: Semiconductor Engineering (2022)
130. LaPedus, M.: Challenges grow in photomask market. In: Chip Design News (2013)
131. Lin, B.J.: Optical lithography with and without NGL for single-digit nanometer nodes. Proc. SPIE. **9426**, 942602 (2015)
132. Burns, G.A., Schoeffel, J.A.: Performance evaluation of the ATEQ CORE-2000 scanning laser reticle writer. Proc. SPIE. **772**, 55–64 (1987)
133. Allen, P.C., et al.: ALTA 4700 system mask patterning performance improvements for X-architecture and wafer electrical performance interchangeability with 50kV E-beam. Proc. SPIE. **5835** (2005)
134. Ljungblad, U., et al.: New laser pattern generator for DUV using a spatial light modulator. Microelectron. Eng. **57-58**, 23–29 (2001)
135. Gil, D., et al.: Parallel maskless optical lithography for prototyping, low-volume production, and research. J. Vac. Sci. Technol. **B20**(6), 2597 (2002)
136. Klosner, M., Jain, K.: Massively parallel, large-area maskless lithography. Appl. Phys. Lett. **84**(15), 2880 (2004)
137. Peng, C., et al.: A high-speed exposure method for digital micromirror device based scanning maskless lithography system. Optik. **185**, 1036–1044 (2019)
138. Martinsson, H., et al.: Current status of optical maskless lithography. J. Microlith. Microfab. Microsys. **4**(1), 011003 (2005)
139. Yao, M.-J., et al.: Application of Sigma7500 pattern generator to X architecture and 45-nm generation mask making. Proc. SPIE. **6607** (2007)
140. Sandstrom, T., et al.: Writing next-generation display photomasks. Proc. SPIE. **9985**, 99851D (2016)
141. Diez, S.: The next generation of Maskless lithography. Proc. SPIE. **9761**, 976102 (2016)

142. Display mask writer (http://www.mycronic.com/)
143. Ge, Q., et al.: Projection micro stereolithography based 3D printing and its applications. Int. J. Extreme Manuf. **2**, 022004 (2020)
144. Chichkov, B.: Two-photon polymerization - high resolution 3D laser technology and its applications. In: Korkin, A., Rosei, F. (eds.) Nanoelectronics and Photonics: from Atoms to Materials, Devices, and Architectures. Springer (2008)
145. Sun, H.-B., Zhang, Y.-L.: Designable 3D nanofabrication by femtosecond laser direct writing. Nano Today. **5**, 435 (2010)
146. Xu, B.-B., et al.: Flexible nanowiring of metal on nonplanar substrates by femtosecond-laser-induced electroless plating. Small. **6**, 1762 (2010)
147. Yang, L., et al.: Multi-material multi-photon 3D laser micro- and nanoprinting. Light: Adv. Manuf. **2**, 17 (2021)
148. Quantum X series 3D Printers (https://www.nanoscribe.com/)
149. Tanaka, T., Sun, H.-B., Kawata, S.: Rapid sub-diffraction-limit laser micro nano processing in a threshold material system. Appl. Phys. Lett. **80**(2), 312 (2002)
150. Haske, W., et al.: 65 nm feature sizes using visible wavelength 3-D multiphoton lithography. Opt. Express. **15**(6), 3426 (2007)
151. Kawata, S., et al.: Finer features for functional microdevices. Nature. **412**, 697–698 (2001)

Chapter 3
Electron Beam Lithography



3.1 Introduction

Electron beam lithography is the most important and also the most versatile nanofabrication technique. Electrons work very much like the photons for nanofabrication. They carry energy and transfer the energy into the energy-sensitive polymer materials called electron resists. The resists exposed by electron beam are developed into surface relief patterns which act as masks for subsequent transfer into the substrate materials.

Electron beam (e-beam) lithography appeared almost around the same time as the optical lithography in the early 1960s. It evolved from scanning electron microscope (SEM). An e-beam lithography system is very much like a SEM. The SEM could evolve to e-beam lithography only because of the discovery of an electron-sensitive polymer material, called polymethylmethacrylate (PMMA) [1]. Exposure of PMMA resist by e-beam is very much like exposure of photoresists by photons. Electron resists are the same group of macromolecular polymers as those photoresists. The exposure of electron resists is the same process as that of photon exposure of photoresists, only that the e-beam lithography can have much higher resolution. As already revealed by Eq. 2.3 in Chap. 2, the resolution of optical lithography is limited by optical wavelength. In order to increase the resolution, the wavelength of light source in optical lithography system has decreased from g-line to i-line, deep UV and extreme UV. Electrons, though are charged particles, are also of wavelike behaviors known as wave particle duality. The electron wavelength can be calculated by Eq. (3.1) [2].

$$\lambda_e = \frac{1.226}{\sqrt{V}}(nm) \tag{3.1}$$

where V is the energy of electrons in electron volt unit. From Eq. (3.1) one can calculate that the wavelength of electrons with 100 eV energy is only 1.2 angstrom.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_3

It is apparent that the higher the electron energy, the shorter the wavelength, which can be hundreds of thousands times shorter than optical wavelength. It is known that high voltage electron transmission microscopes can see individual atoms. This is because of the ultrashort wavelength the high energy electrons have, hence the ultrahigh resolution. In fact, the electron wavelength is not the resolution-limiting factor in e-beam lithography. The factors which limit the achievable resolution in e-beam lithography are various electron aberrations and electron scattering in resists, which will be discussed in the following sections of this chapter.

Nanofabrication realized by e-beam lithography was much earlier than photons did. Because of the fine beam enabled by electron optics, e-beam lithography using PMMA resist immediately demonstrated much higher resolution capability than optical lithography could dream of around the time. In the early 1970s, e-beam lithography was already able to pattern features as small as 60 nm [3]. In the early 1980s when the doom of optical lithography was predicted, e-beam lithography was proposed as the replacement for large-scale IC manufacturing. However, this has never been materialized because of the continuous advances in optical lithography which is inherently high throughput [4]. While patterning by optical lithography is mostly done through photomasks, electron beam can directly expose and pattern on resist, i.e., "direct write." Its high resolution patterning capability has enabled the technology to survive over six decades and has always been a versatile nanoscale patterning tool with wide applications in many fields other than IC manufacturing. With the boom of nanotechnology in the last 30 years, e-beam lithography has become an indispensable nanopatterning tool in research laboratories and universities. Excluding all other factors, the ultimate resolution of e-beam lithography will only be limited by the diameter of electron beam, which has been demonstrated as small as 1.5 nm [5].

There have been attempts to introduce projection e-beam lithography or multiple e-beams for small-volume IC manufacturing. Limited success has been achieved, by which it means that the throughput has been much improved but still not on par with optical projection lithography. The maturity and adoption of EUV optical lithography for sub-10 nm IC manufacturing completely killed the hope for e-beam direct write entering IC production. One thing that e-beam lithography has a firm position in the technology chain of IC mass production is the fabrication of photomasks. Before the advent of laser direct write lithography, e-beam direct write used to be the only tool for making photomasks. Even with today's widely used laser tools, nearly half of all photomasks worldwide are still made by variable shaped e-beam tools. With the increasing complexity of RET (resolution enhancement technology) masks, multi-beam writers became the favored choice, not to mention that e-beam tools are indispensable in EUV mask repair.

This chapter will first introduce the basics of electro-optics, explaining how electrons are focused and deflected. Then the classification of electron beam lithography systems is described, including key specifications. The data formats for e-beam lithography are discussed as part of lithography pattern design process. The exposure process is introduced together with the analysis of electron resists and their characteristics. The difference between e-beam lithography and optical

lithography lies in the fact that electrons are charged particles. They scatter in solid materials, resulting in some special features unique in e-beam lithography, such as the proximity effect caused by electron scattering and how to overcome or correct the effect. Same as the optical lithography, resists play a key part in either high-throughput or high resolution e-beam lithography. Recent developments in e-beam resists and processes are discussed. Finally, progresses in high-throughput e-beam lithography are introduced.

3.2 Principle of Electron Optics

3.2.1 Electron Lens

The core part of an e-beam lithography system is the electron optical system which focuses and deflects an electron beam. To understand the principle of e-beam lithography, one has to first understand the principle of electron optics. Electron optics describes the movement of electrons in an electromagnetic field. The name of electron optics implies that it has many similarities with light optics. It is known that light wave has different propagation speeds in different mediums, resulting in light refraction. Electrons can also change speed and direction of movement in different electromagnetic fields, which is an analogy to refraction. There are, however, major differences between electron optics and light optics [2]:

1. Refractive index in optical mediums is a discontinuous function of space coordinates, i.e., light entering from air to an optical lens undergoes a discontinuous refraction. The refractive index of electron optics is a continuous function of space coordinates because electrical potential always distributes continuously in space.
2. Refractive index of optical mediums is between 1 and 2.5. The refractive index of electron optics depends on the space potential, which can be very large.
3. Geometrical aberrations in light optics can be eliminated by changing arbitrarily the curvature of refractive surface (shape of a lens). However, electrical potential in space cannot be arbitrarily changed. Correction of aberration in electron optics is completely different from that in light optics.
4. Electrons repel each other because they are carrying the same polarity of electrical charge. For a high current density electron beam, there is a space charge effect which can deteriorate the focusing of electron beam. Light optics does not have the space charge effect.

Once understood the relation between electron refraction and space distribution of electrical potential, it becomes easier to understand the principle of electron lens. Figure 3.1 shows the space potential distribution inside two cylindrical electrodes when different voltages are applied to them. It can be seen from observing the potential distribution that the equipotential distribution shapes, spreading from the gap between two cylindrical electrodes, are very similar to the curved surface of an

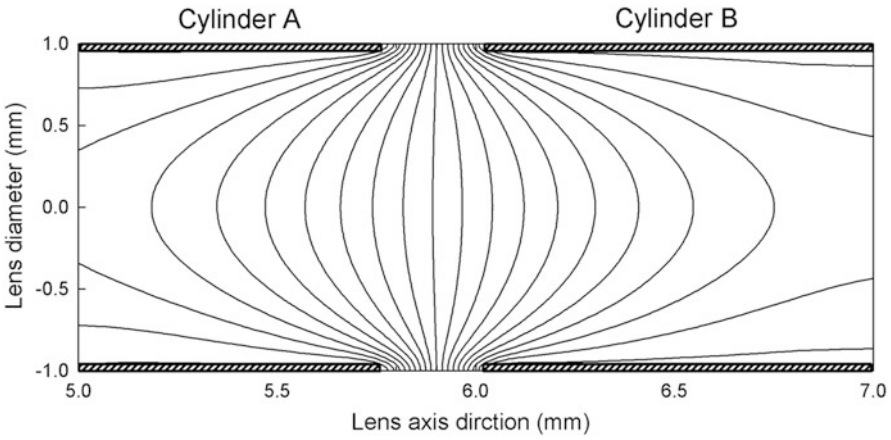


Fig. 3.1 Space potential distribution inside two cylindrical electrodes with different voltages (Cylinder A, 0 volt; Cylinder B, 100 volts)

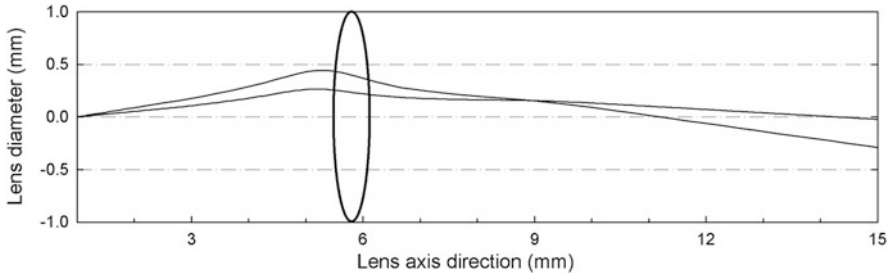


Fig. 3.2 Calculated electron trajectories in the cylindrical electrodes (two trajectories with different incident angles)

optical lens. The difference is that an optical lens only has one fixed curvature of shape, while the curved equipotentials inside the cylindrical electrodes are distributed continuously. An electron coming from one end of the cylindrical electrodes will experience an axial force around the gap region, which can force the electrons from different incident angles and positions to converge toward the cylinder axis. Therefore, the electrical field at the gap region is equivalent to an electron lens. Figure 3.2 shows the two computer-simulated electron trajectories passing through the cylindrical electrode. They come in different angles from the left side of cylinder and converge toward the axis after passing the equivalent lens region.

Focusing by electrical field is called electrostatic lens and focusing by magnetic field is called magnetic lens. The electrodes which generate electrical field distribution can either be cylinders or apertures. They will have lens effect as long as the spatial potential distributions produced are in the form of Fig. 3.1. The structure of magnetic lens is shown in Fig. 3.3. The magnetic field is normally generated by

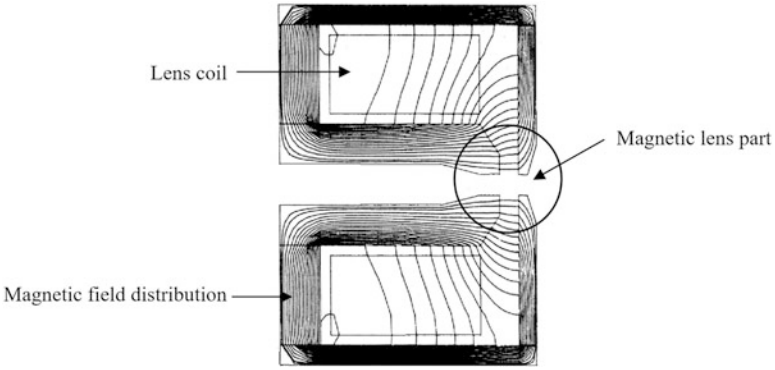


Fig. 3.3 Magnetic lens where the magnetic field is generated by electrical current passing through coils

electrical current flowing through coils which wind over a soft iron core, called pole pieces.

Though the complete theories of electron optics are complex, it can be conceptually understood from the Lorentz force equation as shown in Eq. (3.2). The equation expresses the force exerted to a particle with charge q and speed v in a combined electrostatic field $\boldsymbol{E}$ and magnetic field $\boldsymbol{B}$.

$$F = q(\boldsymbol{E} + v \times \boldsymbol{B}) \tag{3.2}$$

For simplification, assuming the electrostatic field and magnetic field are rotationally symmetric, which is true in most of the electron optical systems, the movement of charged particles can be described by the following equations [6].

For an electrostatic lens with a rotationally symmetric electrostatic field $E(z,r)$, Eq. (3.2) can be expressed as the force experienced by electrons in r and z directions as Eqs. (3.3) and (3.4).

$$m\frac{d^2z}{dt^2} = qE_z = q\frac{\partial V(r,z)}{\partial z} \tag{3.3}$$

$$m\frac{d^2r}{dt^2} = qE_r = q\frac{\partial V(r,z)}{\partial r} \tag{3.4}$$

There is a radial force in r direction to pull electrons toward the axis, as well as an acceleration force in z direction. It is the radial force that focuses electrons into a fine beam. The radial movement of electrons as a function of z can be derived from Eqs. (3.3) and (3.4) as Eq. (3.5), from which the focal length of an electron lens can be derived as in Eq. (3.6).

$$r'' = \frac{(1+r'^2)}{2V(r,z)}\left[\frac{\partial}{\partial r}V(r,z) - r'\frac{\partial}{\partial z}V(r,z)\right]; \quad \text{where} \quad r'' = \frac{d^2r}{dz^2}; \quad r' = \frac{dr}{dz} \tag{3.5}$$

$$\frac{1}{f_i} = \frac{1}{4\sqrt{V_i}}\int \frac{V''(z)}{\sqrt{V(z)}}; \quad \text{where} \quad V''(z) = \frac{d^2V(z)}{dz^2} \tag{3.6}$$

Equations (3.5) and (3.6) reveal three characteristics of focusing charged particles in an electrostatic field.

1. The focusing does not depend on charge or mass of a particle. Whether electrons or ion of different mass will experience the same focusing action.
2. The focusing is directly proportional to the variation of potential distribution along z direction, as indicated by Eq. (3.6). This is the reason that electron focusing only takes place where there is a rapid change of electrical field, which only happens at the gap between two electrodes with different electrical potentials.
3. Equation (3.6) also reveals that high potential *V(r, z)* reduces the focusing action of a lens because the charged particles pass through the lens in high speed.

For a magnetic lens with a rotationally symmetric magnetic field *B(z, r)*,

$$m\frac{d^2z}{dt^2} = qr\frac{d\varphi}{dt}B_r \tag{3.7}$$

$$m\frac{d^2r}{dt^2} = -qr\frac{d\varphi}{dt}B_z \tag{3.8}$$

$$F_\phi = -q\frac{dz}{dt}B_r + q\frac{dr}{dt}B_z, \quad \text{where} \quad F_\phi = \frac{m}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) \tag{3.9}$$

These equations show that electrons not only are focused toward the axis but also rotate around the axis (rotational angle ϕ). $d\phi/dt$ is the rotational speed of electrons. The electron trajectory in the rotationally symmetric magnetic field can be derived from the above equations.

$$\frac{d\phi}{dt} = \frac{q}{2m}B_z; \tag{3.10}$$

$$\frac{d\phi}{dz} = \sqrt{\frac{q}{8mV}}B_z; \tag{3.11}$$

$$\frac{d^2r}{dz^2} = \frac{q}{8mV}rB_z^2 \tag{3.12}$$

For electrostatic lenses, the focusing power depends on how much the difference is of the voltage applied to the lens elements and how rapid is the change of potential. This requires high voltage and very narrow gap between the lens elements, for

example, the two cylinders in Fig. 3.1. Then there is the risk of electrical breakdown if the voltage difference is too great. Therefore, electrostatic lens is in general limited in focusing power. On the other hand, high magnetic field can be generated by high current passing through coils. The heat generated by the high current can be cooled by forced circulation of water or air. In this way, high focusing power can be achieved with a magnetic lens. This is the reason that all the e-beam lithography systems, as well as electron microscopes, use magnetic lenses to focus electron beams. However, magnetic lens cannot be used for focusing an ion beam because of its dependence on charge and mass of ions. A magnetic lens will focus ions of different charges or masses to different position on an image plane. Such a property has been employed for separation of ions in mass spectrometers. For focusing a fine ion beam, electrostatic lenses are generally used.

3.2.2 Electron Source

A fine beam spot focused by an electron optical system is in fact the image of an electron source. To achieve a very small size of beam, the size of its source has to be small. Before the advent of field emission (FE) electron sources, tungsten thermal cathodes were the commonly used electron emission source. Lately, LaB_6 compound was employed as thermal cathode material which performs better than tungsten cathodes. The fundamental drawback of thermionic cathode is their large emission area. In order to reduce the source size, a strong lens in front of the cathode has to be used to focus the electrons emitted from the cathode into a much small cross-over, as illustrated in Fig. 3.4a. This cross-over acts as a virtual source seen by the downstream lenses, so that a fine spot size can be obtained through the focusing system. The problem with this method is that it is difficult to achieve a very small diameter of cross-over because of the Coulomb repulsion of electrons arising from the high concentration of electrons at the cross-over.

Field emission sources have completely overcome this problem. A field emission source does not rely on high temperature heating of the source to generate electron emission. The electrons are directly extracted from the material surface by extremely high electrical field ($>10^8$ V/m). A field emission source is a fine needle. Very high electrical field can be created at the sharp end of needle with relatively low applied voltage [7]. Because electrons only emit at the points where the electrical field exceeds threshold strength, the emission area can be finely controlled and very small. With a field emission source, there is no need to have a cross-over. The virtual source extends to the back of real source, as illustrated in Fig. 3.4b. The virtual source is only a few nanometers. Very high resolution can be achieved using such small virtual source as the object of electron optical imaging system.

Table 3.1 gives a comprehensive list of technical parameters for the four main electron sources which are widely used in various electron beam systems [8]. First, it can be seen that field emission (FE) sources have much small sizes of both the real cathode geometry and the virtual source. Second, their emission current density is

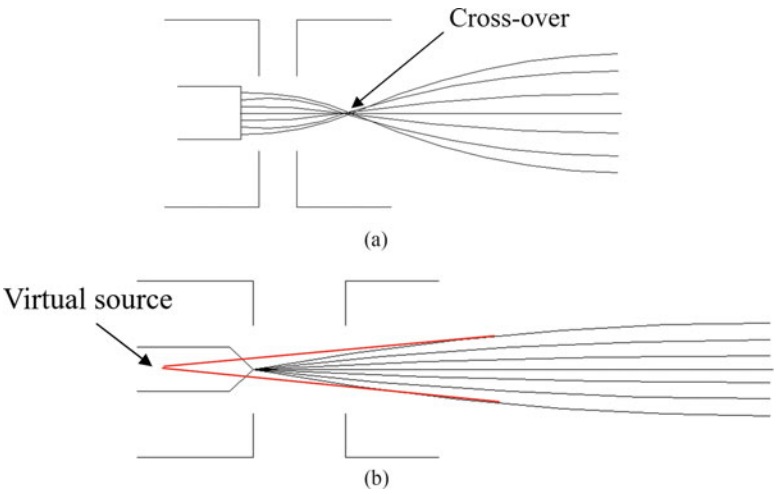


Fig. 3.4 (**a**) Virtual source (the cross-over) in a thermionic cathode, (**b**) virtual source in a field emission cathode

Table 3.1 Comparison of electron emission sources

Emitter type	Thermionic	Thermionic	Cold FE	Schottky FE
Cathode materials	W	LaB6	W	ZrO/W
Operating temperature (K)	2800	1900	300	1800
Cathode radius (μm)	60	10	<0.1	<1
Virtual source radius (nm)	15,000	5000	2.5	15
Emission current density (A/cm^2)	3	30	17,000	5300
Total emission current (μA)	200	80	5	200
Brightness	10^4	10^5	2×10^7	10^7
Maximum probe current (nA)	1000	1000	0.2	10
Energy spread at cathode (eV)	0.59	0.40	0.26	0.31
Energy spread at gun exit (eV)	1.5–2.5	1.3–2.5	0.3–0.7	0.35–0.7
Beam noise (%)	1	1	5–10	1
Emission current drift (%/hr)	0.1	0.2	5	<0.5
Vacuum requirement (Torr)	$\leq 10^{-5}$	$\leq 10^{-6}$	$\leq 10^{-10}$	$\leq 10^{-8}$
Cathode life (h)	200	1000	2000	2000
Cathode regeneration	Not required	Not required	Every 6 to 8 h	Not required
Sensitivity to external influence	Minimal	Minimal	High	Low

several orders of magnitude higher than the thermionic sources. The cold field emission source appears to be the best among the four types of sources. All high-grade electron microscopes nowadays are using cold field emission sources. However, it is rarely used as the electron emission source in an e-beam lithography system. This is because the emission surface of a cold cathode can easily absorb

gaseous atoms, which can affect the electron emission, causing noise and drift of emission current. As seen in Table 3.1, the noise and emission current drift of cold FE source are the highest among the four types of sources. These effects are not serious problems for electron microscope because a flickering of image does not matter very much. But they are unacceptable to e-beam lithography application. E-beam lithography relies on stable emission current to deliver consistent exposure dose. Any emission current drift or noise can seriously deteriorate the quality of resist exposure. This has been the reason to develop thermal field emission cathodes for e-beam lithography systems. The thermal field emission cathode, also called Schottky cathode, is to heat up the conventional field emission cathode. At high temperature, absorption of gaseous atoms at cathode emission surface is greatly reduced, resulting in stable working of the cathode. As seen in the table, Schottky cathode has the most characteristics of a field emission source but much lower in beam noise and emission current drift.

3.2.3 Aberrations

A small size of source does not guarantee a final small size of beam. An electromagnetic lens system must have as small as possible aberrations in order to achieve a fine beam size. Similar to light optics, an electron optical system also has geometrical aberrations, such as spherical aberration, coma, field curvature, astigmatism and distortion, and nongeometrical aberration such as chromatic aberration. As aforementioned, the fundamental difference between electron optics and light optics is that the refractive index in an optical lens is a discontinuous function of space coordinates, i.e., light entering from air to an optical lens undergoes a discontinuous refraction, while the refractive index of an electromagnetic lens is a function of electromagnetic field distribution, which is a continuous function of space coordinates. The geometrical aberrations in light optics can be eliminated by changing arbitrarily the curvature of refractive surface (geometrical shape of an optical lens). However, electromagnetic field in space cannot be arbitrarily changed. An optical imaging system may consist of a hundred pieces of lenses to correct all the aberrations. An electromagnetic imaging system has only a few lenses. A typical e-beam lithography column is schematically illustrated in Fig. 3.5 [9]. The electron gun is at the top, which consists of an electron source and some lenses to initially focus the electrons, as illustrated in Fig. 3.4. The condenser lens further focuses the beam. In conjunction with the beam blanker and selectable aperture, the beam can be switched on or off (pass or not pass the aperture). The deflectors perform the beam scanning function. The final objective lens focuses the beam onto a substrate surface.

Due to aberrations, an electron beam spot on axis (off-axis deflection is not considered here) at the substrate (image plane) is not simply a demagnified electron source but includes other aberration components, as expressed in Eq. 3.13 [10].

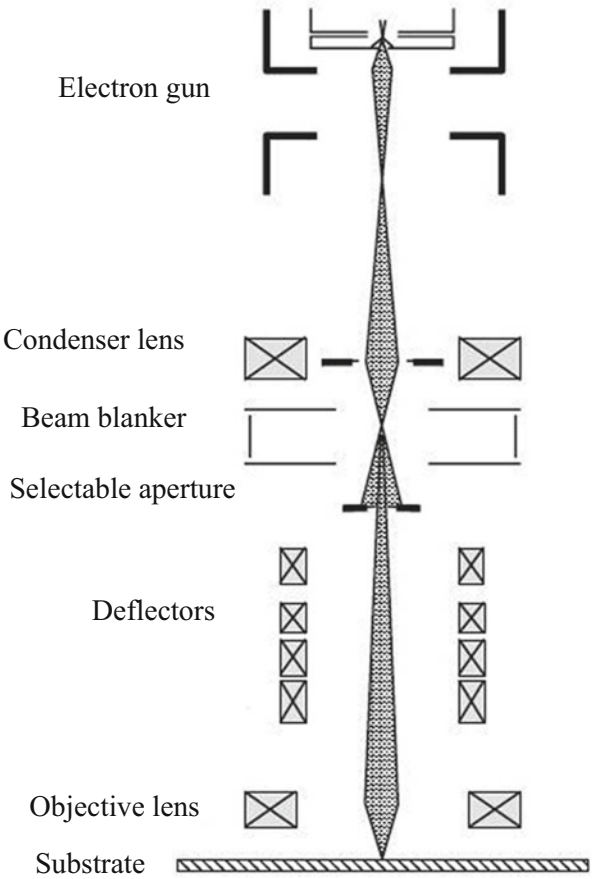


Fig. 3.5 Schematic of complete electron optical system in an e-beam lithography column. (Reprint from [9] with permission)

Beam spot diameter:

$$d = \sqrt{d_g^2 + d_s^2 + d_c^2 + d_d^2} \tag{3.13}$$

where:

Beam spot without aberration: $d_g = \frac{d_v}{M}$ (d_v, virtual source diameter; M, demagnification).

Spherical aberration: $d_s = \frac{1}{2} C_s \alpha^3$ (C_s, spherical aberration coefficient; α, beam convergence angle).

Chromatic aberration: $d_c = C_c \alpha \frac{\Delta V}{V}$ (C_c, chromatic aberration coefficient; ΔV, beam energy spread; V, beam energy).

Diffraction limit: $d_d = 0.6 \frac{\lambda}{\alpha}$ (λ (nm), electron wavelength $\lambda = \frac{1.2}{\sqrt{V}}$nm).

Spherical aberration is caused by large angle of beam. Electrons entering a lens with different angles will be focused differently, as the trajectories shown in Fig. 3.2. Although an electron optical system can be designed to minimize the spherical aberration coefficient *Cs*, the most effective way of reducing the spherical aberration is to reduce the angle α, because the aberration is proportional to the cubic power of α. The α can be reduced by using an aperture to prevent wide angular electrons from entering the objective lens. The selectable aperture in Fig. 3.5 performs exactly this function. For the objective lens, the aperture is the source. Its size has great influence on the final spot size on the substrate. An e-beam lithography system is normally equipped with a series of different sizes of apertures. They are selectable and exchangeable to suit different requirements on beam current and resolution. It is obvious that for ultimate high resolution e-beam lithography, the smallest aperture should be selected, albeit the usable beam current is also greatly reduced. Unfortunately, reduction of convergent angle will cause increase of electron diffraction aberration. Although the wavelength of high energy electrons is extremely small, as indicated in Eq. 3.1, the diffraction can be noticeable if the aperture is also extremely small. Obviously, increasing beam energy can significantly reduce the diffraction effect.

Chromatic aberration is caused by energy spread of emitted electrons, and by the instability of applied voltage in the case of electrostatic lenses or instability of current in the case of magnetic lenses. While the fluctuation of external power supply to an electromagnetic lens system can be well controlled to minimum, the energy spread of an emission source is, however, beyond anyone's control. Therefore, the chromatic aberration is always there. The energy spread can be exasperated if there is a beam cross-over in the case of thermal emission source. At the cross-over, high concentration of electrons creates significant repulsion among the charged particles, causing increase of the energy spread. The evidence is displayed in Table 3.1, where it indicates that electrons leaving a thermionic cathode have energy spread of 0.59 eV but increase significantly to 1.5–2.5 eV at the electron gun exit, because of the cross-over formed in front of the cathode. Apparently, the cross-over has made the energy spread much worse. Not only the cross-over at the electron gun side, any cross-over along the axis of an electromagnetic lens system can create additional energy spread, hence the chromatic aberration. Therefore, an advanced electron optical system will try to avoid any cross-over. For example, the GEMINI column developed at ZEISS does not have any cross-over along the path from a field emission source. The only cross-over is the focus point at the substrate surface where the beam lands [8].

Another aberration which is unique in electron optics is caused by space charge effect. As mentioned earlier, electrons at a cross-over will experience Coulomb repulsion force. The repulsions among electrons not only increase the energy spread but also increase the diameter of beam, resulting in defocus and radial blur of the beam [11]. The higher the emission current, the greater the space charge effect, hence more difficult to achieve small beam diameter in an electromagnetic lens system. E-beam lithography has much greater demand on beam current than a SEM, because a resist can only be exposed above a threshold exposure dose, whereas the

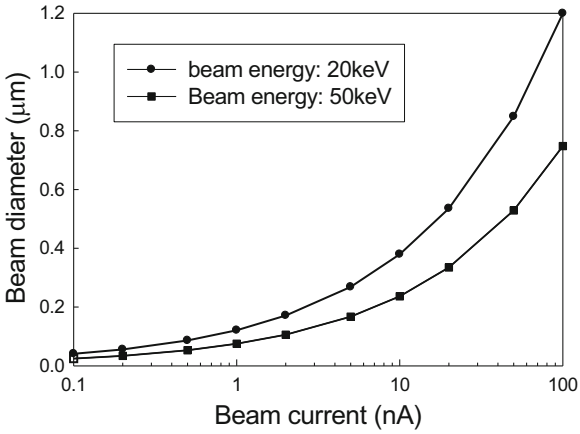


Fig. 3.6 Beam diameter vs. beam current

exposure dose is the product of beam current and beam dwelling time. High beam current can reduce the beam dwelling time, resulting in high throughput in e-beam lithography. However, high beam current makes the focusing of electron beam more difficult. That is partly the reason that a SEM can have much small size of electron beam than can an e-beam lithography system. The other reason is that SEM can use a cold field emission source while e-beam lithography system cannot, as explained previously. A modest scanning transmission electron microscope (STEM) in the 1970s had already been able to provide a 0.5-nm-diameter electron beam. The current at this beam size was 1pA (10^{-12}A) [12]. In contrast, an e-beam lithography system rarely operates below 100pA beam current. The normal beam current is 0.1 nA ~ 100 nA.

The only way to reduce the space charge effect without compromising the beam current is to increase the beam energy, in other words, operating at high acceleration voltage in an e-beam lithography system. High beam energy can also minimize the chromatic aberration caused by beam energy spread, because both the space charge induced beam enlargement and the chromatic aberration are inversely proportional to the beam energy. This is partly the reason why high-end e-beam lithography systems always operate at 100 keV beam energy. The relation between beam current and beam diameter is best demonstrated in Fig. 3.6, where beam diameters vs. beam current at 20 keV and 50 keV beam energies are compared [13]. Therefore, either smaller beam current or higher beam energy, or both, is preferred when high resolution e-beam lithography is required.

Besides the abovementioned aberrations which are associated with beam focusing, there are also aberrations related to beam deflection. They are mainly distortion and astigmatism which increase in cubic power proportion with the distance of deflection from the center axis. An e-beam lithography system normally has scan fields from a few hundreds of micrometers to a few millimeters. The beam spot is no longer small and circular at the corners of deflected field. They become bigger and elliptical, and a square scanning field becomes distorted [14]. For example, a previously reported prototype e-beam lithography system had achieved beam spot

diameter of 2.7 nm on axis, but the beam spot deteriorated to 6.5 nm at the edge of 500 μm deflection field [9].

The above analysis indicates that a high resolution e-beam lithography requires high beam energy, small scanning field, low beam current, and a stable environment. These are the prerequisites to enable low aberrations and fine focusing of a beam.

3.3 Electron Beam Lithography Systems

3.3.1 Basic Configuration

Though the basic components of an e-beam lithography system have been illustrated in Fig. 3.5, a practical system is much more complex, as shown in Fig. 3.7, which is generally referred as electron column. An electron column consists of the following subsystems:

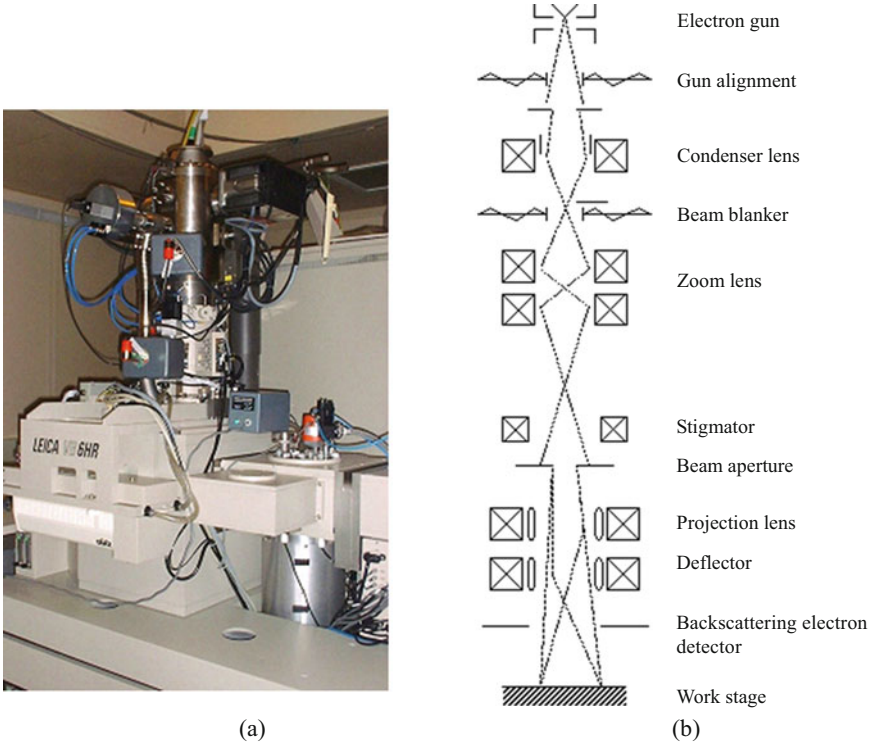


Fig. 3.7 (**a**) Photograph of an e-beam lithography system (VB6-HR from Leica Microsystems) and (**b**) schematic of complete electron optical system in an e-beam lithography column

1. Electron gun. The electron gun has been described in detail in the last section. What needs to be mentioned here is that electron gun is always at very high voltage. Electron beam in an e-beam lithography system is normally accelerated to a kinetic energy between 10 keV and 100 keV. The acceleration voltage is always applied to electron gun, i.e., the gun is kept at negative high voltage while the exposure stage is at ground potential, so that a high voltage difference is established between the gun and the stage with which electrons emitted from the gun are accelerated to land at the stage. If the cathode is field emission cathode, a separate vacuum system is needed at the gun section in order to maintain a high vacuum for field emission of electrons.
2. Electron gun alignment system. The whole electron column is assembled section by section from each part of subsystems. The assembled column can be as 1 meter high. Any slight error in machining or assembling of sections can lead to misalignment between the cathode tip in the gun and the last stage of lens aperture. A deflection system at the electron gun section is normally installed to collimate the electron beam if an alignment is required.
3. Condenser lens. It has the same function as the condenser lens in an optical lithography system, to converge the emitted electrons from a cathode so that they arrive at the exposure surface in maximum numbers.
4. Beam blanking. This is to switch on and off the electron beam so that the beam can only pass through the column to reach substrate when an exposure is needed. It functions like a shutter in a camera, by deflecting the beam off the electron optical axis so that it cannot pass through a central aperture.
5. Zoom lens. This is for adjusting the focal plane, including dynamic focusing of electron beam.
6. Stigmator. Astigmatism occurs when focusing of a beam is different in x and y directions, resulting in an elliptical shape of beam instead of a circular beam. It is mostly caused by the errors in mechanical machining of the lens components. Stigmators are generally composed of multipole lenses, which can focus a beam laterally from different directions.
7. Apertures. Several apertures are used in an electron column. One of the important functions of an aperture is to limit the angle of electron beam spread. The main aberration of an electron lens is the spherical aberration which is directly proportional to the cubic power of beam spread angle. Limiting the angle is an effective way to improve electron optical resolution. However, the use of small aperture can reduce beam current by cutting off large-angle electrons. In practice, the size of aperture can be adjusted (using a series of apertures of different diameter). A smaller aperture is used when high resolution is required and a larger aperture is used when high beam current at low resolution is needed. The position of aperture can be finely tuned to ensure the center of aperture coincides exactly with the axis of electron column.
8. Projection (objective) lens. This is the final lens of an electron column, which focuses the electron beam coming out the aperture to form the final beam spot at the exposure surface.

9. Deflectors. Electron beam deflection and scanning are realized by deflectors. The deflectors can be placed before, after, or inside the projection lens. A deflector can be electrostatic or magnetic. Magnetic deflector is a coil which generates magnetic field by electrical current flow. A magnetic deflector has small aberration and distortion but slow deflection speed, whereas electrostatic deflection is of high speed but larger aberration. An e-beam lithography system is usually equipped with two sets of deflection system, responsible for main field and sub-field deflection. An exposure field (main field) is normally divided into many sub-fields. Main field requires large deflection angle but at low speed, whereby a magnetic deflector can be used. Sub-field deflection is of small angle and high speed, which can be carried out by an electrostatic deflector.
10. Electron detector and work stage. An electron detector is equipped at the exit of final lens to detect the backscattered electrons from the substrate. This is to detect the alignment mark when aligned exposure is required. Unlike a SEM which detects secondary emission electrons, backscattered electrons are less efficient for imaging. A work stage is where the substrate is mounted. The stage moves in synchronization with beam deflection to enable exposure of large-size substrate.

Other peripheral components include vacuum systems, high voltage power source, and pattern generator which converts a design pattern data into electrical signals to drive beam scanning deflectors.

3.3.2 Key Specifications

When choosing an e-beam lithography system, one should pay attention to the following key specifications:

1. Minimum spot size, which is the indication of system resolution. It directly influences the minimum size of exposed features.
2. Acceleration voltage, which is the indication of electron beam energy. A system with high acceleration voltage also has higher resolution because of smaller aberration (mentioned earlier) and smaller electron proximity effect (discussed later). High energy beam can also expose thicker resist layer.
3. Beam current. A system with larger beam current can produce higher speed of lithography, i.e., less time is needed to expose a pattern at the same exposure dose than a low beam current system. However, the exposure speed is limited by the scanning speed of a system, or the response speed of beam deflectors. Besides, a high current beam always has large spot size. In general, a small beam current is used for high resolution lithography and large beam current for large pattern exposure which gives high throughput.
4. Scanning speed. High scanning speed results in shorter exposure time. Scanning speed is expressed in frequency. The maximum frequency of current vector scan

systems is 50 MHz, while the maximum frequency for raster scan systems is up to 500 MHz. Vector scan and raster scan will be discussed in the next section.
5. Scanning field. For a large scanning field, most patterns can be exposed within the field, which can avoid field stitching errors. However, large field has low scanning speed and high deflection aberration.
6. Other specifications include the accuracy of work stage, accuracy of overlay between different exposure levels, accuracy of field stitching, etc.

3.3.3 *Vector and Raster Scanning*

The difference between vector scan and raster scan is that in vector scan electron beam only scans the exposure patterns (Fig. 3.8a), whereas in raster scan the electron beam scans the whole field all the time, but the beam blanker is switched on only where there are patterns (Fig. 3.8b). Vector scan systems only scan when there are patterns. The electron beam is a Gaussian beam, or a circular beam. It is commonly called Gaussian beam because the current distribution in the circular beam spot is close to a Gaussian distribution function. Vector scan systems generally have high resolution but low scan speed comparing to raster scan systems. The fastest vector scan system so far is only 50 MHz scanning speed. However, the high resolution capability ($\leq$3 nm beam spot size) makes these systems as powerful tools for nanoscale lithography.

In a practical raster scan system, the electron beam only scans in one direction, because of the limit of scanning field size, while the work stage moves in other direction. They are combined to obtain a large field size. Raster scan e-beam lithography was exclusively used for making photomasks. These systems, represented by MEBES series, were typically working at 10 keV and minimum feature size around 350 nm. Because of the low resolution, they were used for mask making until the 130 nm node IC technology. The masks for 130 nm node onward were exclusively patterned by 50 keV variable shaped e-beam systems (see Sect. 3.7.1).

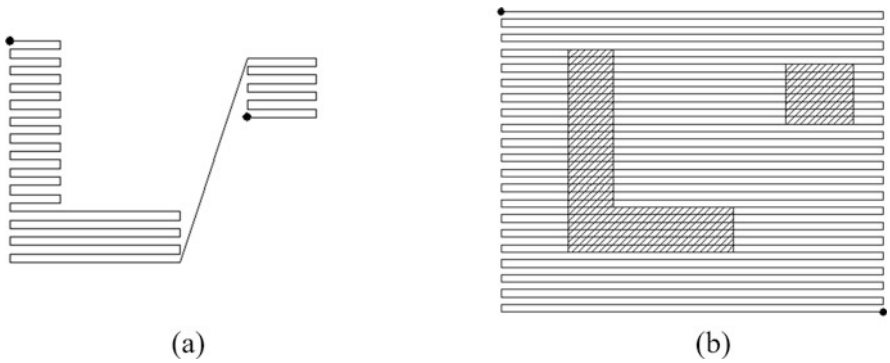


Fig. 3.8 Schematic illustration of (**a**) vector scan and (**b**) raster scan

3.3.4 Pattern Fragmentation

The deflector in an e-beam lithography system has limited deflection range usually from a few hundred micrometers of sub-field to a few millimeters of main field, which is adjustable. Patterns bigger than the field size have to be fragmented. Figure 3.9 illustrates that a big pattern is divided into a number of fragments to fit into different deflection fields. The exposure of whole pattern is completed by stitching together these fragments. There are two types of stitching, sub-field stitching within a main field and main field stitching between movements of work stage. Modern commercial e-beam lithography systems are equipped with laser-controlled work stage which has very high accuracy of movement. Both sub-field stitching and main field stitching can be controlled within 25 nm accuracy [15]. One must be aware of the field size and stitching accuracy of an e-beam lithography system before designing very fine patterns. During the design process, patterns should be grouped as much as possible within an exposure field. For those patterns bigger than the exposure field, care should be taken to avoid putting critical parts of the pattern at the field boundary.

Apart from setting the field size, pixel size and step size also need to be specified. Note the difference between pixel size and step size, illustrated in Fig. 3.10. In short, electron beam stops at each pixel and moves to next pixel in steps. The step size can be equal to pixel size or a few times bigger than pixel size. By adjusting the pixel size and step size, the way a pattern is filled by electron beam can be adjusted.

As most of e-beam lithography systems can only scan a pattern in x and y directions in a Cartesian coordinate system, any patterns, no matter how complex they are, have to be fragmented to basic shapes of rectangles, trapeziums, or parallelograms. Electron beam scans in raster scan mode inside each of the basic shapes and moves from one basic shape to another in vector scan mode, as schematically shown Fig. 3.8a. The fragmentation is much straightforward if the exposure pattern is a rectangular shape or can be decomposed into a number of rectangular shapes. If a pattern is a curved shape, the curved boundary has to be

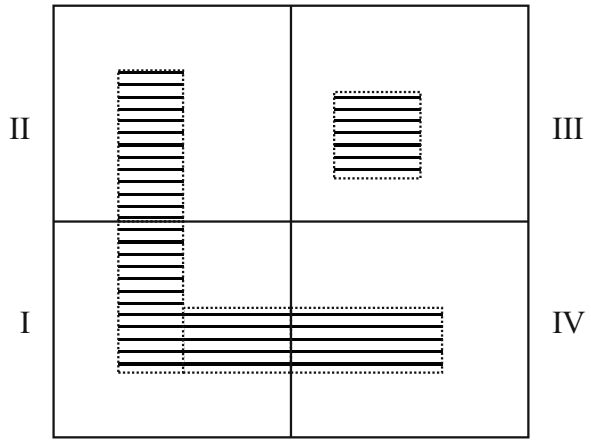


Fig. 3.9 Pattern fragmentation according to exposure field (I, II, III, and VI are the exposure fields covered by the pattern)

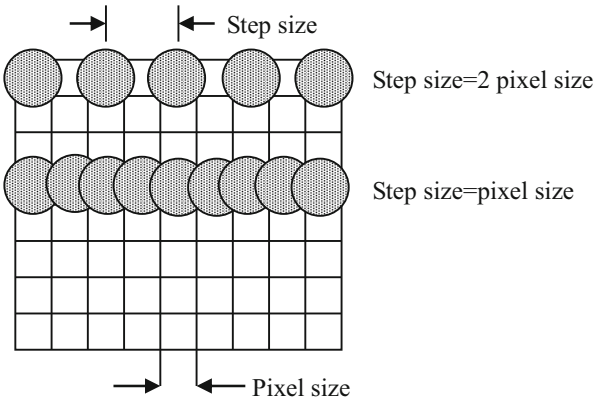


Fig. 3.10 Definitions of step size and pixel size of scanning exposure

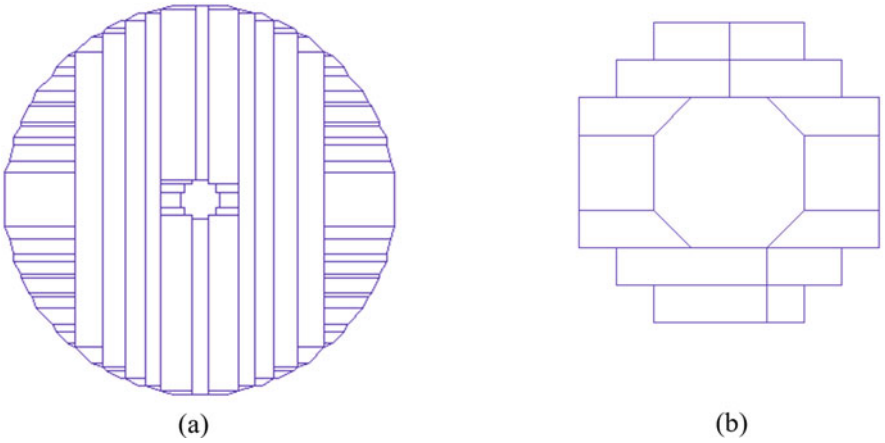


Fig. 3.11 (**a**) Fragmented ring shape pattern with outer diameter 1 μm and (**b**) fragmented ring shape pattern with outer diameter 80 nm

approximated by a large number of rectangles or trapeziums, resulting in a large number of fragmentations. Figure 3.11 is an example of fragmentation on a ring shape pattern. If the curved shape is big, the rectangle and trapezium fragmentation can generally approximate the curved boundary in good precision, as shown in Fig. 3.11a. However, if the curved shape itself is very small and comparable to the resolution limit of the e-beam system, approximation by rectangles can be very crude because of the limit of pixel size and step size, as shown in Fig. 3.11b.

It indicates that Cartesian scan systems cannot obtain high precision curved patterns. This is not a major problem in most e-beam lithography applications because most of the e-beam systems are for making electronic devices which are mostly rectangular patterns. Curved patterns are encountered mainly in making

optical elements such as waveguides. There is a fixed beam moving stage (FBMS) technique which can pattern any curve-shaped patterns and without stitching errors [16]. As the name implies, FBMS is a scanning mode that the electron beam is stationary and an exposure pattern is formed by continuously moving the work stage. In this way, any pattern shapes can be exposed without the need for pattern fragmentation and without any stitching errors, such as long curved optical waveguides [17].

3.3.5 Commercial E-Beam Lithography Systems

For high-quality sub-100 nm patterning, a commercial e-beam lithography system is necessary. There are used to be a number of providers of commercial systems. Over the years, the market has been consolidated into two major players and one small player. The two major players are Raith represented by its top grade e-beam lithography system EBPG Plus [18] and JEOL represented by its top grade system JBX-9500FS [19]. Raith used to be a middle range e-beam lithography system provider, while Vistec Lithography (originally Leica Cambridge) was in the top league in direct competition with JEOL. In 2013, Raith acquired the part of Gaussian beam system of Vistec to become the top league. Vistec is now focusing on variable shaped beam systems for advanced mask making [20]. The small player is NanoBeam which is specialized in high resolution e-beam direct write systems. Its nB5 system is capable of patterning 6 nm features with good uniformity over a large area [21]. All these top grade commercial e-beam lithography systems operate at 100 keV beam energy and with scanning speed of at least 50 MHz, and all of them are capable of below 10 nm patterning resolution.

The aforementioned professional e-beam lithography systems are of high performance and multifunction but also very expensive with multimillion US$ price tags. Not many universities or research institutes can afford such expensive systems. On the other hand, most of scientific researches do not need a top grade e-beam lithography system. There is another large group of Gaussian beam vector scan systems which are modified from scanning electron microscopes. It functions both as a SEM and an e-beam lithography system. Raith is the main supplier of such combined systems. The company still has this main product line represented by Raith150 which operates at maximum 30 keV beam energy and 20 MHz speed and can pattern sub-10 nm feature size.

If one has already got a SEM, there is an even cheaper solution, that is, to modify the existing SEM into an e-beam lithography machine. As mentioned earlier, a SEM is fundamentally the same as an e-beam lithography system, except that a SEM does not have a pattern generator, a laser-controlled work stage, and a beam blanker. Therefore it can be converted into a lithography tool once these kits are installed. Raith, apart from selling high-end e-beam lithography systems, also provides the ELPHY upgrade kits, known as “lithography attachments.” The ELPHY kits enable virtually any SEM to be converted into an e-beam lithography tool without

sacrificing any of its original functionality [22]. It must be pointed out that the functionalities of these SEM converted systems are very limited because [23]:

1. A SEM normally does not have a beam blanking function. Without beam blanking, the electron beam will not be switched off when moving from one exposure pattern to the next, resulting in some unwanted exposure features in resist. Nowadays, many SEM manufacturers have provided the option to equip a beam blanker.
2. SEM works in raster scan mode, whereas SEM converted e-beam lithography system must be able to work in vector scan. The deflection coil in a SEM cannot completely adapt to vector scan mode. As seen in Fig. 3.8a, vector scan mode is needed when jumping from one pattern to another. A SEM deflection coil is not designed for such a jump. When a pulse signal is applied to the coil, it needs a fairly long settling time to become stable. For example, JEOL840 SEM needs 10 μs settle time and Hitachi S800 SEM has settling time up to 1 ms, which can severely hamper the lithography speed.
3. The work stage movement in a SEM is normally manual. Even if there is a step motor, the accuracy of stage movement is not better than 1–5 μm. When the exposure pattern is larger than the scan field, field stitching becomes a serious problem. It can be improved by adding a laser-controlled work stage. However, a laser-controlled stage is normally quite expensive.

In general, the main limitation of a SEM converted e-beam system is its inability for large area and high-throughput e-beam lithography. Nevertheless, it is still a useful tool for scientific research wherein only small-size samples with a few devices or structures may be needed. Longer exposure time is not a serious problem and no quantity of devices is required. In such circumstance, the SEM converted systems become very attractive because they are much less expensive.

3.4 Scattering and Proximity Effect

To realize nanoscale lithography or direct nanofabrication with electron beam, a small beam size is necessary but not enough. Electrons have to interact with a polymer resist, so that the energy carried by the electrons is transferred to the polymer material, causing changes of molecular chains in a polymer. The energy transfer process takes place during collisions between the electrons and the atoms in the polymer. The process is called scattering, which can lead to a much larger area of energy deposition than the original size of beam landing area, creating the so-called proximity effect. The proximity effect is the major obstacle in the process of achieving high resolution nanofabrication by e-beam lithography.

3.4.1 Electron Scattering

High energy electrons penetrating into a solid material will collide with atoms and the electrons surrounding an atom. As we have learned from general physics, two things can happen when two rigid balls collide with each other. If a small ball collides with a massive ball, the small ball will only change the direction of movement without losing any energy it carries. If a small ball collides with another small ball of same mass, the small ball will transfer half its energy to the other ball. The similar scenario happens in electron scattering in solid. The massive balls are the atoms and the small balls are electrons. There are two types of scattering taking place in the solid: elastic and inelastic scatterings. Collision of incident electrons with atoms in the material is an elastic scattering. As the mass of an electron is at least 1/1840 of atomic mass, the collision is an extremely small ball with an enormous ball. An electron undergoing elastic scattering does not lose its energy but simply change its direction. Collision of incident electrons with the electrons surrounding the atoms of solid material is inelastic scattering, during which the incident electron will pass part of its energy to the electrons of solid material. It is this part of transferred energy that causes chain reaction of resist polymer material, resulting in the exposure of resist. Therefore, the electron scattering process in a solid material can be pictured as this: electrons from an incident electron beam undergo continuous elastic scattering by material atoms and continuously change their flying directions. In between two elastic scatterings, there are inelastic scatterings with material electrons whereby the beam energy is gradually deposited into the material.

The rigid ball collision is only an analogy to describe the scattering process taking place during electron beam exposure of resist. Electrons and atoms are not rigid balls. Their collisions are through the interaction of energy field around the atoms and electrons. The interaction radius of energy field is described by scattering cross section. For elastic scattering there is Rutherford's single electron differential scattering cross-section model. It describes that if a solid material consists of N different elements, the differential scattering cross section per solid angle of the ith element can be expressed as Eq. (3.14).

$$\frac{d\sigma_i}{d\Omega} = \frac{e^2 Z_i (Z_i + 1)}{64(\varepsilon_0 \pi E)^2} \left(\frac{E + m_0 c^2}{E + 2m_0 c^2} \right)^2 \frac{1}{\left(\sin^2 \left(\frac{\theta}{2} \right) + \alpha_i^2 \right)^2} \tag{3.14}$$

Integrating Eq. (3.14) over the entire solid angle results in the total elastic scattering cross section, as shown in Eq. (3.15).

$$\sigma_i = \frac{e^2 Z_i (Z_i + 1)}{64(\varepsilon_0 \pi E)^2} \left(\frac{E + m_0 c^2}{E + 2m_0 c^2} \right)^2 \frac{1}{\alpha_i^2 (1 + \alpha_i^2)} \tag{3.15}$$

where σi is the scattering cross section of ith element in unit of m^2, $d\Omega$ is the differential solid angle, e is electron charge, Z_i is the atomic number of ith element,

ε_0 is the permittivity of free space, E is energy of incident electrons in unit of eV, θ is the deflection angle, and α_i is the screen parameter.

Inelastic scattering can be described by Bethe's continuous slowing down approximation model [24], as shown in Eq. (3.16).

$$\frac{dE}{ds} = -\frac{N_A \rho e^2}{8\pi\varepsilon_0^2} \sum_i \left(\frac{C_i Z_i}{A_i}\right) \frac{1}{E} \ln\left(\frac{aE}{I_i}\right) \tag{3.16}$$

where dE/ds is the rate of energy loss of unit travel length (eV/m), N_A is the Avogadro constant, ρ is the density of solid material (kg/m^3), A_i is the atomic weight of ith element, a is a constant (1.1658), C_i is the weight fraction of the ith element in the material, and I_i is the mean excitation energy (eV).

The above two equations can be used to calculate the electron scattering and energy deposition in solid material. The collision of electrons with atoms in solid is a random process. Which atom the electron collides with and which direction it is deflected are all random. Therefore, the scattered electron trajectories are also random, which can be modelled by Monte Carlo simulation [25, 26]. Monte Carlo method is a numerical technique for solving mathematic equations by a random process. Here it is not to use the method to solve mathematic equations but to borrow its name to reflect the random nature of electron scattering process. Figure 3.12a shows two electron trajectories (primary electron A and B), when they travel inside a resist layer on top of substrate material. The resist is PMMA of 0.1 μm thickness and the substrate is silicon. It was simulated by a Monte Carlo simulation program MOCASEL [27]. Each dot along a trajectory represents a scattering event. One can see that the position, direction, and free travel distance of each scattering event are totally random. Yet the probability of a scattering event taking place and the energy loss of a primary electron are governed by the physical law of electron-atom collisions. Although each individual scattering is random, simulation of a large number of electron scatterings will generate the general picture of the extent of energy deposition due to the collision and loss of energy from the primary electrons. Figure 3.12b plots out 100 electron trajectories. The electron energy is 20 keV with 1 μm resist layer on top of a silicon substrate. This plot gives a rough idea how deep and how wide the electrons travel inside resist and substrate. Most of electrons scatter forward into substrate but some electrons scatter back into resist layer. These forward scatterings and backward scatterings (backscattering) are the cause of broadening of exposure area in resist.

To generate statistically correct and reliable results, the number of electrons to be simulated has to be between 50,000 and 100,000. This number does not represent the real number of electrons in an electron beam, but only represents all the possible scattering directions of a single electron. Its energy deposition is represented by the so-called point spread function (PSF) which is the deposited energy distribution in resist from single point of incidence. Once the PSF is obtained, the energy deposition by a real electron beam can be calculated by convoluting the PSF with a beam spread function, for example, a Gaussian function for a Gaussian beam. The typical PSF is

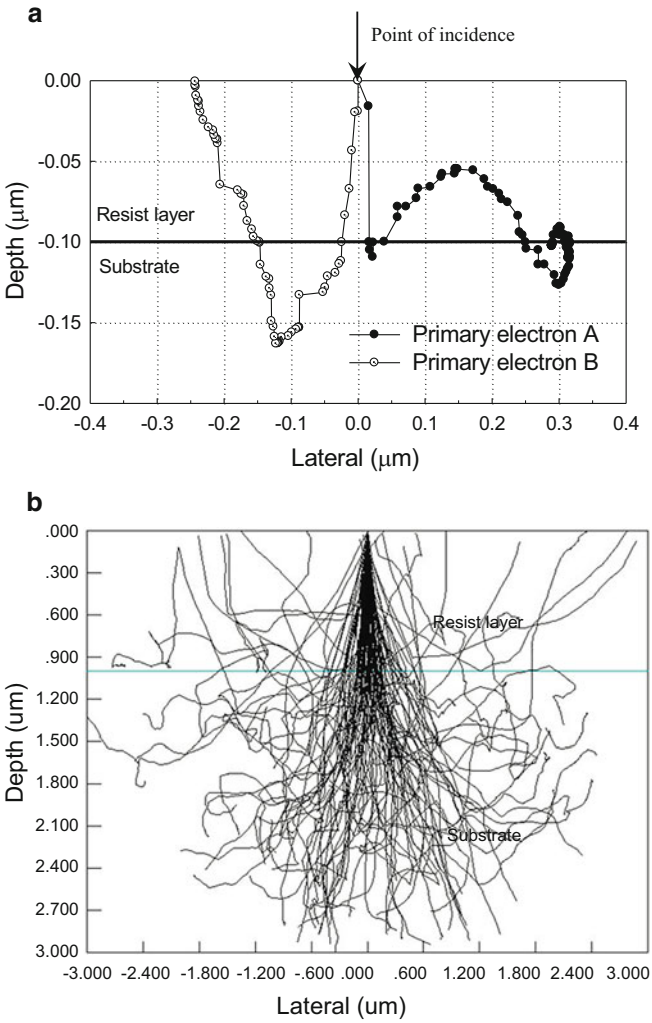


Fig. 3.12 (**a**) Plots of two electron trajectories showing random scattering events, (**b**) plots of 100 electron trajectories showing the extent of electron scattering

shown in Fig. 3.13, where forward scattering part and backscattering part of energy spread are indicated. Figure 3.13 also compares the PSFs at 10% and 90% depth of resist. It clearly displays a significant increase of broadening in deposited energy by forward scattering, as well as accumulated energy by backscattering, as electrons go deeper.

There are three factors that strongly influence the electron scatterings:

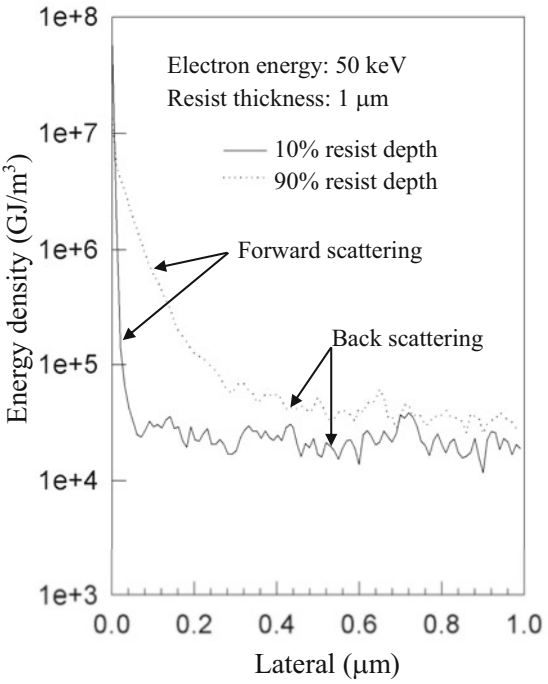


Fig. 3.13 Point spread functions of electron energy distribution at different depths of resist

1. Electron energy

Higher energy of primary electron has much smaller scattering angle, therefore, less broadening of the beam due to small-angle forward scattering. Figure 3.14 compares the PSFs of 50 keV, 70 keV, and 100 keV electron energies. The higher the electron energy is, the narrower the PSF becomes. This is another proof that high energy electron beam is preferred in e-beam lithography, in addition to previously mentioned benefits in lowering chromatic and space charge aberrations.

2. Density of substrate material

More electrons will scatter back from substrate if the substrate material is of higher density. Figure 3.15 compares the PSFs for silicon and GaAs substrates. Electron energies are both at 50 keV and the resist thickness is the same as 0.5 μm. However, the deposited energy level due to backscattering electrons is significantly higher for GaAs substrate than for silicon substrate, because the density of GaAs is 2.3 times higher than that of silicon. Even if the substrate is a light material such as silicon, the backscattering can still be significant if there is a thin film of heavy metal, such as gold or tungsten, coated at the interface between silicon and resist layer. Therefore, it is more difficult to achieve high resolution e-beam lithography on high density substrate.

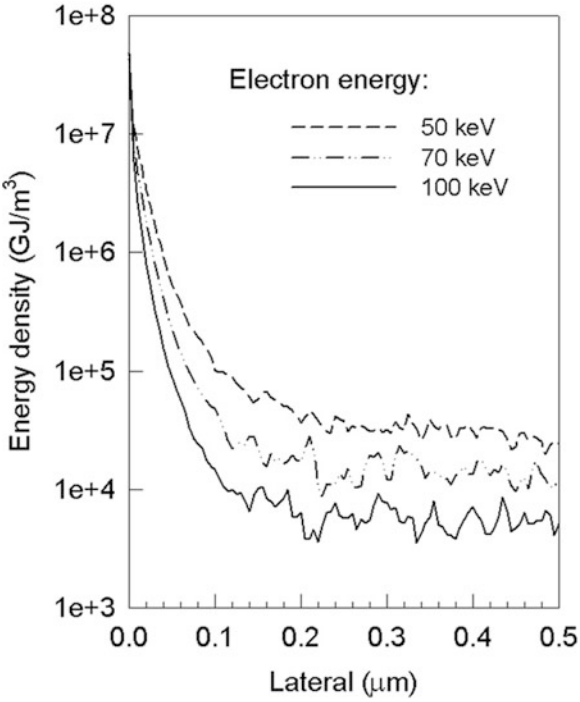


Fig. 3.14 Comparison of PSF at different electron energies (silicon substrate, 0.5 μm resist thickness)

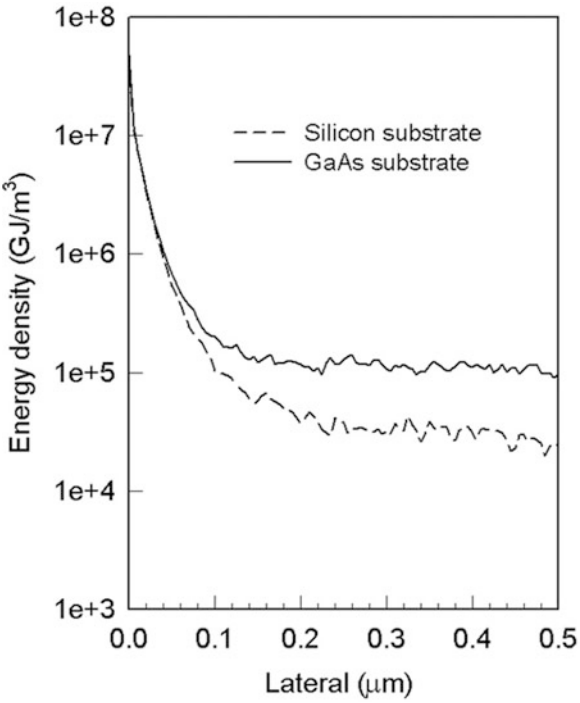


Fig. 3.15 Comparison of PSF for different substrate materials (50 keV beam energy, 0.5 μm resist thickness)

3. Resist thickness

This has been apparent from Figs. 3.12b and 3.13 that electrons scatter wider forwardly as they go deeper. Therefore, for high resolution e-beam lithography, thin resist layer is preferred.

3.4.2 *Proximity Effect and Correction*

Like the optical proximity effect which is caused by light diffraction, the direct consequence of electron scattering in e-beam lithography is the so-called electron proximity effect. Because the deposited energy from scattering electrons can extend to adjacent areas, dense features will become overexposed, and very isolated feature will be underexposed, creating the so-called intershape and intrashape proximity effect. An example of proximity effect from computer simulation is shown in Fig. 3.16. It shows the cross-section profile of PMMA resist of 1 μm thickness, exposed by 20 keV electron beam. The pattern is 100 nm line grating with 500 nm pitch. PMMA is a positive tone resist. As described in Sect. 2.3.2 of Chap. 2, resist is removed after exposure and development for positive resist. The ideal profile is

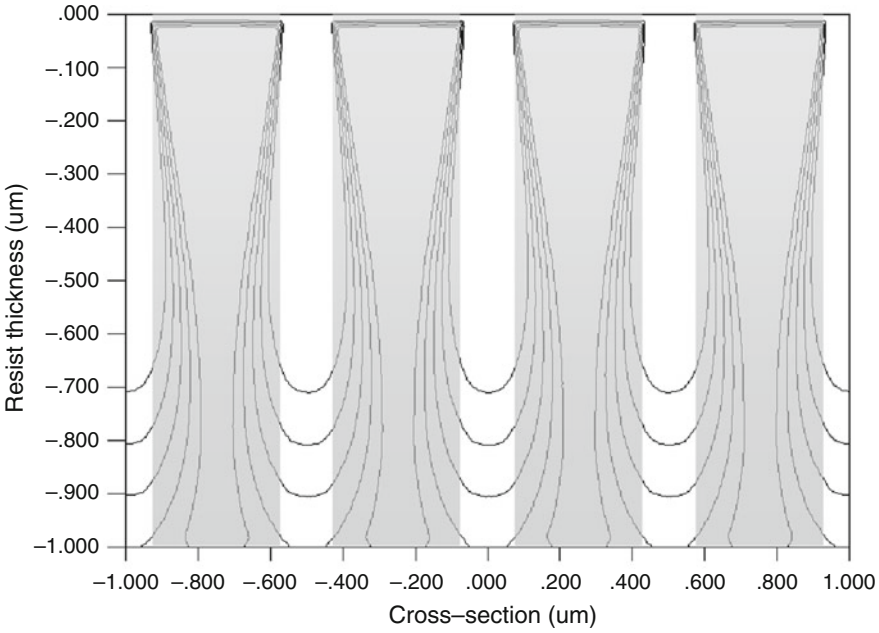


Fig. 3.16 Simulated resist cross-section profiles for 20 keV electron beam exposure of 1 μm PMMA resist layer on silicon substrate: the shaded area represents the ideal resist profile, and the contours represent the actual resist profiles at exposure doses from 80μC/cm^2 to 120μC/cm^2 with 10μC/cm^2 increment

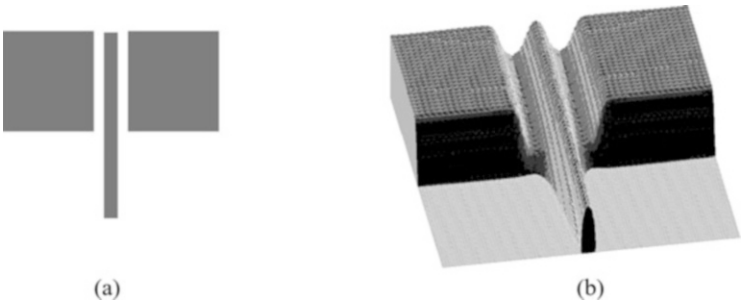


Fig. 3.17 (**a**) Exposure patterns with same exposure dose, (**b**) computer-simulated 3D resist profile without proximity correction

shown by the shaded area which is unexposed resist with 100 nm line openings. The actual resist profile is shown by the contours which represent the evolution of resist cross section at different exposure doses. The higher the exposure dose, the deeper it goes until it reaches the full depth of 1 μm. It is apparent that the actual resist profile (contour lines) is significantly deviated from the ideal profile (shaded area), because of the proximity effect. The result manifests the effect of electron forward scattering, which becomes wider as the electrons go deeper in the resist.

For high resolution e-beam lithography, particularly when dealing with irregular pattern size and density, correction of proximity effect is absolutely necessary. Proximity correction has always been part of e-beam lithography development and can be dated back to more than 40 years ago [28]. Many research results have been published on this subject. In this section, the three principal correction methods are briefly introduced. They are dose correction method, pattern size compensation method, and background exposure compensation method.

1. Dose correction is the most widely used correction technique and also a technique giving the best correction results. As previously shown, electron scatterings create nonuniform distribution of electron energy in resist, which extends beyond the exposed pattern area and is the root cause of proximity effect. The principle of dose correction is to adjust the exposure dose at different parts of a pattern so that all patterns receive the same exposure energy. For example, Fig. 3.17a shows a group of patterns. They all receive the same exposure dose, represented by the same shading color. Because of the electron scattering and proximity effect, the line between the two big pads cannot be resolved, as the simulated 3D resist profile shown in Fig. 3.17b. If the patterns are given different exposure doses, as shown in Fig. 3.18a where the light color shading represents lower exposure dose, the line in the middle can be fully resolved as shown by the simulated 3D resist profile in Fig. 3.18b. In this case, the big pads receive less exposure dose, and the middle line receives different exposure doses at different sections depending on whether it is between the pads or outside the pads.

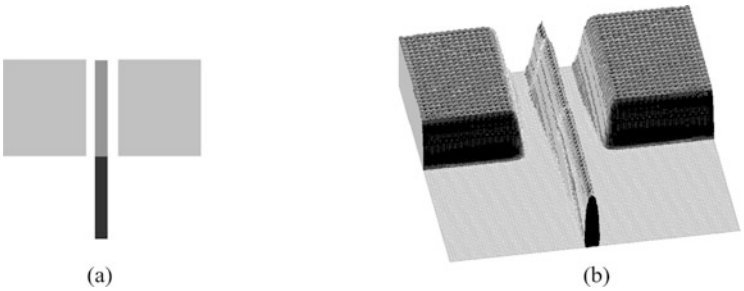


Fig. 3.18 (**a**) Exposure patterns with different exposure doses at different locations, (**b**) computer-simulated 3D resist profile with proximity correction

The above example is a very simple case, for which one can intuitively alter the exposure dose on different parts of the patterns. For much complicated pattern designs, proximity correction has to be done by intensive computation. A self-consistent method has been developed to find the correct dose allocation for a design pattern [28]. The method first assumes an initial exposure dose applying to all patterns. Energy distribution of all the patterns, including contributions from neighboring patterns due to electron scattering, is calculated. Naturally, the distribution will not be uniform. The next step is to adjust the exposure dose on various regions of the exposure area and various parts of each pattern based on the known energy deposition. Now the exposure doses are no longer uniform. Calculation of energy is carried out again according to the new exposure doses. The uniformity of energy distribution should be improved, but may not be good enough. Another round of dose adjustment and energy calculation may be needed. The cycle is repeated until the deviation of energy distribution between two calculations is within an error limit (tolerance). It is then considered a self-consistent solution has been reached and the iteration can be terminated. The final result will be a map of exposure dose which has different value at different part of exposure patterns. Theoretically, the self-consistent method requires calculating the electron scattering effect for all the patterns. If the pattern number is N, there will be N sets of equation to solve. The amount of computing is enormous and may not be necessary. As we know that the electron scattering range is limited around the point of incidence, the self-consistent method can be simplified by zoning the exposure area, i.e., the calculation of scattering and energy deposition is only carried out in certain zones where electron scattering effect is significant [29].

The procedure of getting correct exposure doses by self-consistent solution is of direct physical meaning, hence called a physics-based correction method. Its solution is also relatively accurate. However, to find self-consistent solutions for an IC design where there are millions of pattern shapes, the simulation procedure becomes very complicated, and the computation is very time-consuming. There is another simple geometry fraction method, which gives much faster proximity correction, though the correction result is relatively coarse. The difference of the two methods is

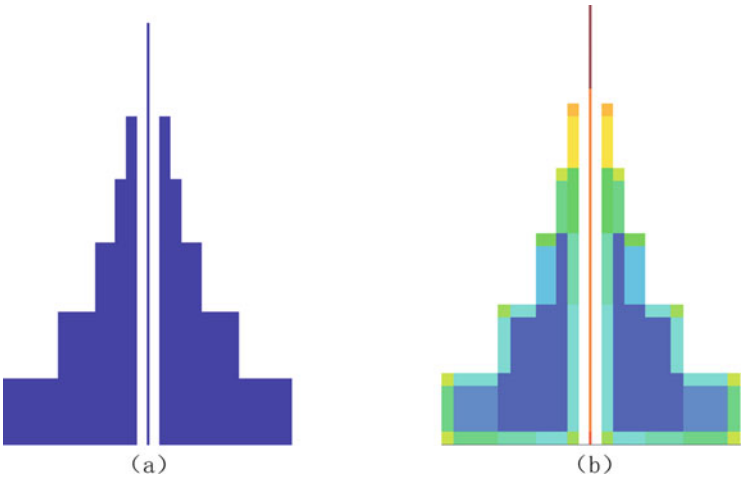


Fig. 3.19 An exposure pattern (**a**) without geometry fraction; (**b**) with geometry fraction

that the physics-based correction fractures a pattern according to the self-consistent calculation of electron energy at every point of the pattern. So the fractured patterns are irregular. Geometry fraction method first divides a pattern into regular sub-shapes according to the radius of electron scattering range and then calculates the electron energy deposition from which appropriate doses are assigned to each sub-shape [30]. It is a faster proximity correction method because there are no iterative computations. The electron energy in all patterns is only calculated once.

The above proximity correction requires complicated computer algorithms. There are a number of commercial proximity correction software available. Usually the proximity correction software comes as part of software tool sets when an e-beam lithography system is purchased, for example, the NanoPECS proximity effect correction package supplied by Raith. These commercial software tools can process any pattern designs with reasonably accurate correction results. Fig. 3.18 is a simple example of geometry fraction, where the exposure doses are assigned manually. An example of proximity correction by computer is shown in Fig. 3.19. The original pattern with single exposure dose (single color) is fractured and assigned with different exposure doses to different parts of the pattern (different colors).

2. The proximity correction can be done by resizing the pattern geometry, which is the pattern size compensation method. The resizing, also called biasing, is best suitable for simple and highly repeated patterns, such as lines and squares. For a group of repeated lines, the lines in the middle of pattern area can be resized narrower (negatively biased) than nominal design to compensate the higher than average energy contributed from neighboring lines, or the lines at the outer rim of pattern area can be resized wider (positively biased) to compensate the lower than average energy due to electron scattering. The overall result is a relatively uniform distribution of energy across the whole pattern area. The amount of

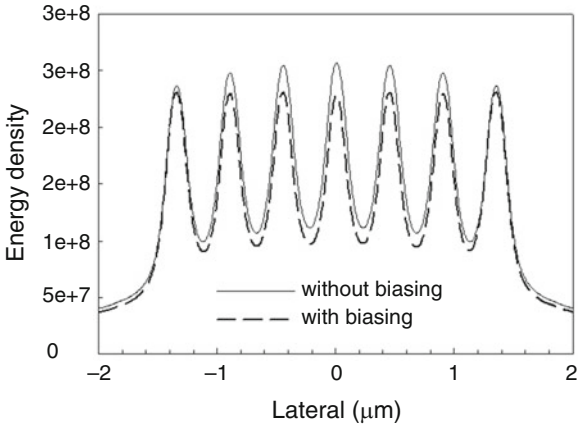


Fig. 3.20 Comparison of pattern biasing for proximity effect correction (150 nm lines and 450 nm pitch, exposed with 20 keV beam energy into 0.5 μm resist on silicon)

biasing can be determined empirically or from simulation of electron energy distribution. An example is shown in Fig. 3.20. It shows deposited energy for a group of seven lines with linewidth of 150 nm and pitch of 450 nm and exposed by 20 keV electron beam into 0.5-μm-thick resist on silicon. If all lines are exposed at equal width, the lines in the middle receive higher energy (solid line plot) because electron scatterings from both sides have contributed energy to the middle lines. If the middle lines are negatively biased by 10 nm to 25 nm, all the lines receive more or less the same energy (dashed line plot).

3. Proximity correction can be done using background exposure to compensate the nonuniformity of electron energy distribution. As aforementioned, electron scattering always causes excess of energy at dense pattern region and less energy at sparse pattern region. If a second exposure is carried out at all non-pattern regions (background), the pattern density for the second exposure is exactly reversed. The originally sparse pattern region now becomes dense pattern region and receives more energy. Meanwhile, the original dense pattern region where there are less blank areas now receives less energy. The superimposition of the two exposure can well balance the energy distribution so that it becomes uniform everywhere. Figure 3.21a shows the energy distribution of a group of seven lines (150 nm width and 450 nm pitch). The first exposure produces an unbalanced energy profile, with the lines in the middle receive higher energy. If a second background exposure, with energy profile shown by the dashed line plot in Fig. 3.21a, is added to the first exposure, the overall energy distribution is well balanced, as the energy profile plot shown in Fig. 3.21b. The background exposure method has a name called "GHOST" [31]. It is simple and does not need special proximity correction software and does not need to calculate the electron energy distribution. All it needs is to have a second exposure of reverse polarity of the original pattern design. The second exposure can be carried out with a defocused electron beam and at relatively low exposure dose, with only the

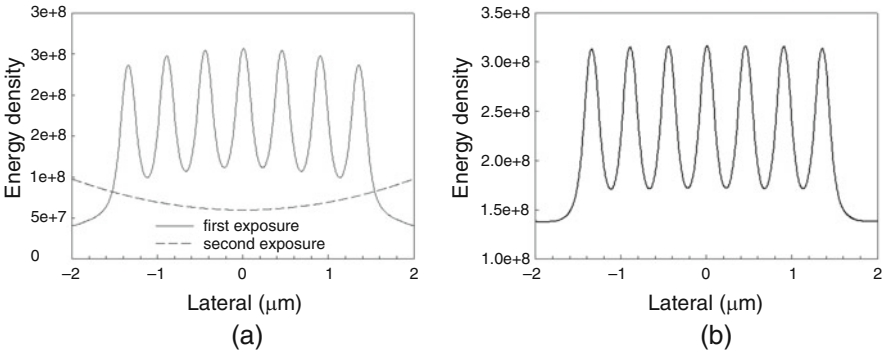


Fig. 3.21 (**a**) Energy profiles of first and second exposures, (**b**) superimposed energy profile

purpose of compensating the nonuniformity of energy. The drawbacks of this method are increased pattern data volume, increased total exposure time, and deterioration of image contrast due to the background exposure. This technique is only practical in a raster scan e-beam system. A raster scan system scans the whole exposure field all the time. The beam is only switched on where there is a pattern shape to be exposed. Therefore, the background exposure is just another scan with beam switched on in opposite polarity. For a vector scan system, the beam only scans over a pattern area. The background exposure requires another set of design data with all the background patterns and the exposure time can be exceedingly long.

All the above proximity correction techniques require additional process and special software tools. A simple and effective measure to reduce the proximity effect is to increase beam energy as well as to decrease resist thickness. High beam energy can greatly reduce the beam expansion caused by forward scattering and reduce localized backscattering. As forward scattering is wider with electrons going deeper, a thin resist layer can also minimize the effect of beam expansion. An example is shown in Fig. 3.22, which is a direct comparison with Fig. 3.16. The pattern design and resist thickness are the same in both cases. The only difference is the beam energy for the exposure. While in Fig. 3.16 the resist is exposed at 20 keV, the beam energy in Fig. 3.22 is 100 keV. No proximity effect is seen for 100 keV exposure and resist profile is almost the same as the ideal profile (shaded area).

All the above discussion on electron scattering and proximity effect has favored using high energy electron beam. High beam energy results in narrower forward scattering and lower backscattering. Although the higher, the better, too high voltage makes the design and construction of electron optical column very difficult. Besides, electron energy above ~120 keV can damage silicon lattice and semiconductor devices. Therefore, it has been concluded that 100 keV is the best suitable energy for e-beam lithography [32]. However, one must be aware that high beam energy requires high exposure doses. The exposure dose increases linearly with the increase of beam energy, because majority of primary electrons have gone deep into substrate and very few of them contribute to the exposure of resist.

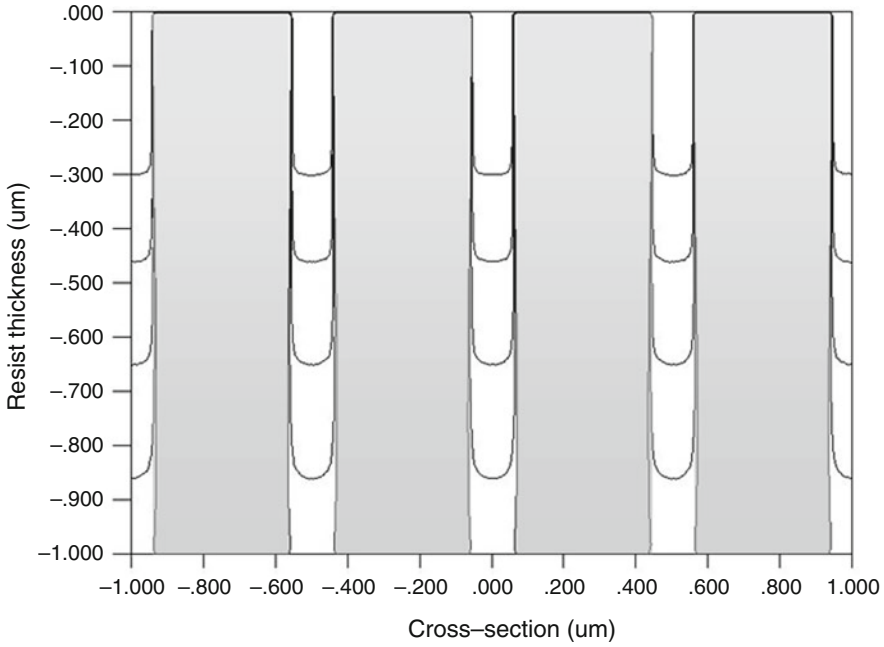


Fig. 3.22 Simulated resist cross-section profiles, to compare with those in Fig. 3.16, for 100 keV electron beam exposure of 1 μm PMMA resist layer on silicon substrate: the shaded area represents the ideal resist profile, and the contours represent the actual resist profiles at exposure doses from 150μC/cm^2 to 350μC/cm^2 with 50μC/cm^2 increment

3.4.3 Effect of Secondary Electrons

So far the discussion about electron scattering and proximity effect has only been focused on primary electrons. Although the major contribution of proximity effect comes from scattering of primary electrons, secondary electrons can play a vital role when e-beam lithography is conducted close to the beam diameter, namely, around 10 nm or below. The secondary electrons are electrons knocked off from the atoms of solid material by high energy primary electrons. While traveling inside the resist, primary electrons transfer their energy to the electrons around the atoms of materials. These secondary electrons acquiring up to kilovolts of energy become mobile and can travel a short distance. For example, electrons of 400 eV energy can travel about 12 nm and 200 eV electrons can travel about 5 nm distance [33]. They scatter the same way as primary electrons, only in a short distance, but always in directions perpendicular to the path of primary electrons, and leave their energy along the path of scattering. The consequence of the scatterings of these "fast" secondary electrons is a broadened energy deposition in resist. Figure 3.23 gives a comparison of point spread functions (PSFs) with and without contributions from secondary electrons by

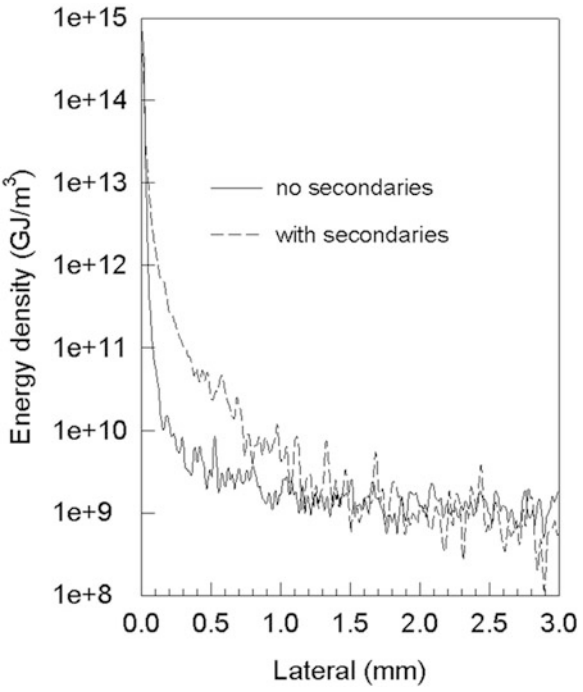


Fig. 3.23 Comparison of PSFs with and without secondary electron contribution

Monte Carlo simulation. It is apparent that secondary electrons made a noticeable contribution to the low energy part of PSF.

It was concluded decades ago that on average the secondary electrons scatter about 10 nm distance around the main path of a primary electron, which limits the e-beam lithography resolution to no better than 10 nm, no matter how small the e-beam spot size is [34]. However, this conclusion is not completely true, because e-beam lithography depends not only on electron energy deposition but also on resist material and process. The latter can to some extent suppress the energy broadening effect caused by secondary electrons and enable lithography resolution to go below 10 nm. The e-beam resist process will be discussed in a later section of this chapter.

3.4.4 Low Energy E-Beam Lithography

Although high beam energy has been favored in e-beam lithography, its rather inefficient use of energy, i.e., the need of high exposure dose and possible damage to substrate material or to buried structures, has prompted investigation on very low energy exposure of resists. At very low end of beam energy (0.5–5 keV), primary electrons cannot travel very far. Figure 3.24 compares the electron scattering trajectories at 10 keV energy (Fig. 3.24a) and 3 keV energy (Fig. 3.24b). The resist thickness is 100 nm on a silicon substrate. At 10 keV, most of the electrons have

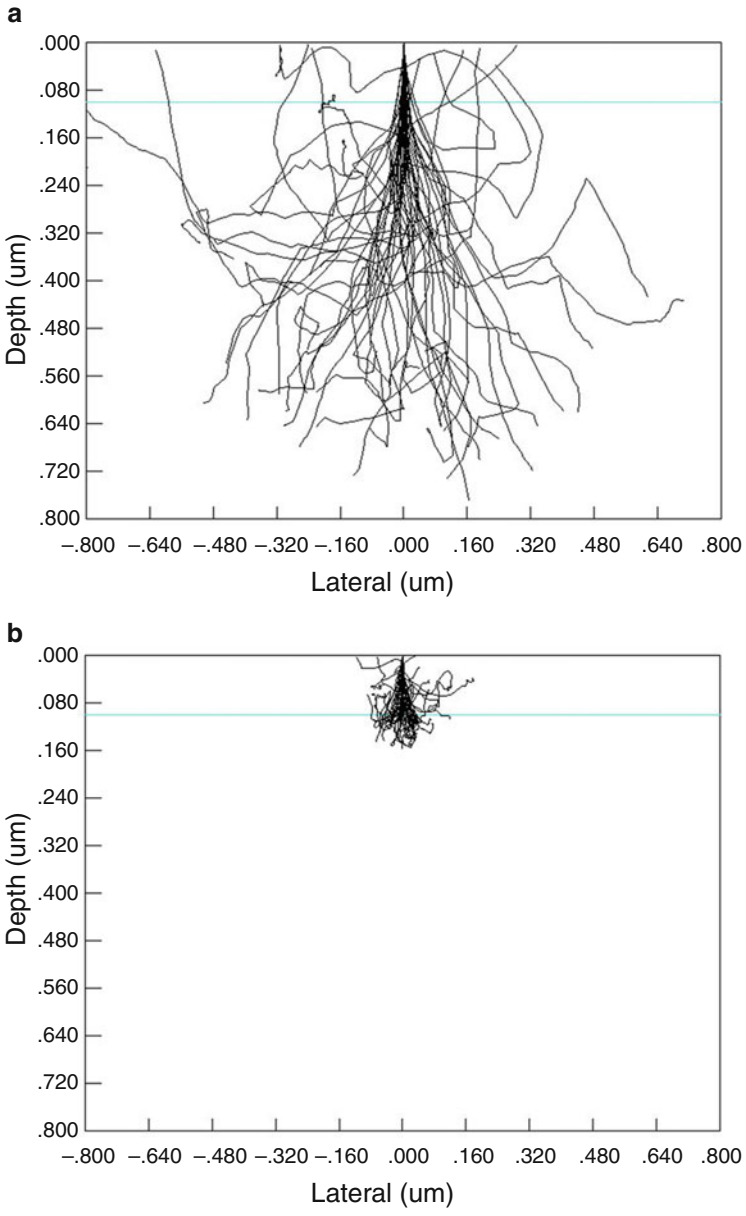


Fig. 3.24 Comparison of electron scattering at 10 keV energy (**a**) and 3 keV energy (**b**)

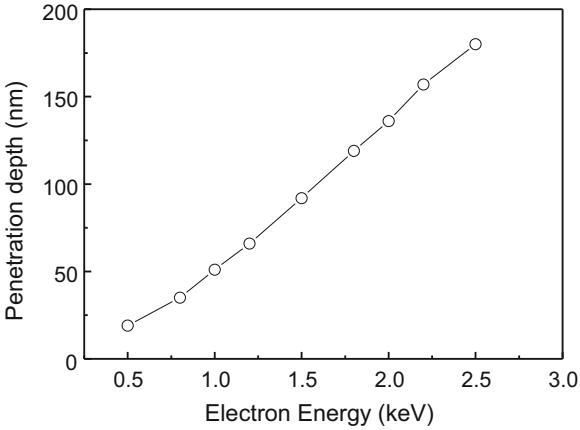


Fig. 3.25 Electron penetration depth at different energies

gone into substrate and there are backscattered electrons as well. The higher the beam energy, the more electrons would end up deep in substrate. Two problems can arise: one is substrate heating and another is substrate charging. For nonconductive substrate materials, such as quartz and glass which are commonly used for making photomasks, both heating and charging exist. Charging problem can be avoided by using a metallic underlayer between resist and nonconductive substrate [35]. Resist heating due to deposition of large amount of electron energy in a substrate of poor thermal conductivity can cause change of resist sensitivity and severe proximity effect [36], particularly in e-beam lithography of photomasks [37], and there is no effective way to overcome the heating problem.

With low energy electrons, there would be little energy deposited into substrate if a correct combination of resist thickness and beam energy is chosen. Figure 3.25 shows experimentally measured penetration depth in resist at different beam energy. The resist was ZEP-520 positive e-beam resist. A series of large square patterns were exposed by an e-beam with different energies. The resist was then developed and the remaining thickness was measured, from which the penetration depth was worked out [38].

Because the electron penetration is shallow at low energy, it brings two advantages: One is small range of forward scattering. If the resist thickness is just about the penetration depth of low energy electrons, then there will be no backscattered electrons, hence no backscattering-related proximity effect. For example, sub-100 nm lines were produced in 66-nm-thick PMMA layer at 2 keV beam energy, and the deviation of linewidth between dense pattern region and sparse pattern region was found to be less than 10 nm [39]. Another advantage of shallow penetration of electrons at low energy e-beam is the high concentration of energy deposition in resist layer so that fewer electrons are needed for resist exposure. Resists become much more sensitive to low energy electrons. Figure 3.26 is the measured resist sensitivity vs. electron energy [38]. The resist (ZEP-520) was 180 nm thick and electron energy ranged from 3 keV to 20 keV (the electron

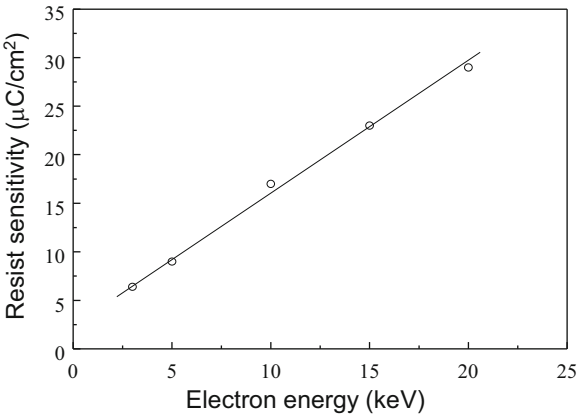


Fig. 3.26 Resist sensitivity vs. electron energy

penetration depth is over 180 nm at 3 keV energy as shown in Fig. 3.25). The resist sensitivity is defined as the exposure dose needed to enable the 180-nm-thick resist layer to be fully developed. As seen from the plot, the decrease of resist sensitivity (increase of required exposure dose) is linearly proportional to the increase of beam energy. In fact, low energy electrons are much more efficient at exposing a resist polymer, because the majority of polymer chain scission or cross-linking happens during interaction with electrons of below 100 eV energy. For high energy e-beam lithography, it is the low energy secondary electrons, excited by high energy primary electrons, which are responsible for the exposure of resist [33].

The shallow penetration of low energy e-beam lithography is also a drawback, because only very thin resist can be used for lithography. The resist thickness should normally be below 100 nm, depending on the beam energy. As is known from previous chapter, thin resist is less resistant to reactive ion etching, making resist pattern transfer to substrate more difficult. In this case, multilayer resist structure has to be used, whereby the top thin layer resist acts as an imaging layer [40]. Another drawback of low energy e-beam lithography is the difficulty to achieve fine size of beam. As has been described in previous section, both chromatic aberration and space charge effect become severe at low beam energy. The performance of a conventional e-beam lithography system falls rapidly below 6 keV acceleration voltage. The only way to maintain relative focusing quality of the electron optical column is to introduce a retarding field after the objective lens, to slow down high energy e-beam to a low energy before landing on the substrate [41].

A more promising low energy e-beam lithography technology was proposed in the late 1990s, which was based on projection of low energy electrons using a stencil mask [42]. The resolution of low energy e-beam projection lithography (LEEPL) is no longer limited by the beam spot size but by the resolution of stencil mask. A 70 nm feature dimension has been achieved by LEEPL at 2 keV electron energy on 50-nm-thick ZEP-520 resist with no proximity effect [43]. However, the requirement of a stencil mask stopped the LEEPL from further development due to the same

reason as for X-ray lithography or other projection-based e-beam lithography. Presently, the only commercial system that offers low energy e-beam lithography is the Raith150 Two which can set beam energy from 100 eV to 30 keV, although low energy e-beam lithography can also be implemented simply using a SEM converted e-beam lithography tool for scientific research purpose.

3.5 Resist Materials and Processes

The energy of electrons deposited in a polymer resist will have to be converted into molecular chain reaction in a resist to eventually form a lithographic pattern. On the one hand, a small beam size and small scattering range are not necessarily translated into a small size of resist features. On the other hand, a poor resist material and non-optimum resist process can certainly result in poor resolution and much enhanced proximity effect. Resist materials and processes play a vital role in achieving the ultimate resolution capability of e-beam lithography. The basics of polymer resists and specifics for optical lithography have been mentioned in Sects. 2.3.2 and 2.7.4 of Chap. 2. The present section deals with specifics of e-beam resists and the factors influencing the final achievable e-beam lithography resolution.

3.5.1 Sensitivity

The most important properties of a resist are sensitivity and contrast, which are directly linked to the resolution capability of the resist. The sensitivity and contrast are derived from a resist development curve which typically looks like the plots in Fig. 3.27 (more information about the resist development curve can be found in

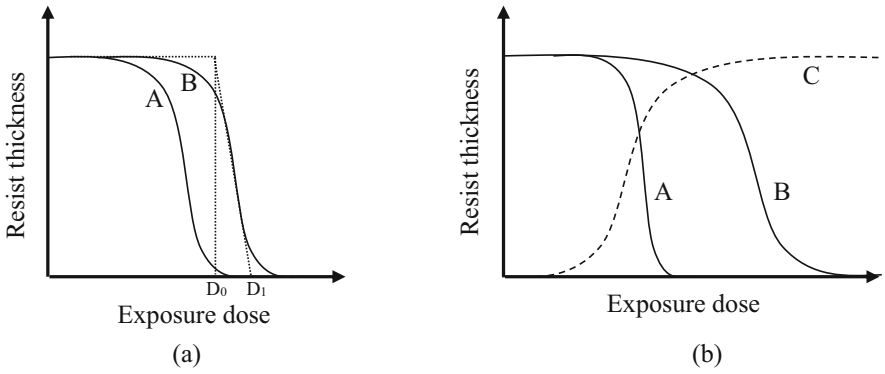


Fig. 3.27 Illustration of resist development curves: (**a**) resist A is of higher sensitivity than resist B; (**b**) resist A is of higher contrast than resist B, and resist C is negative resist

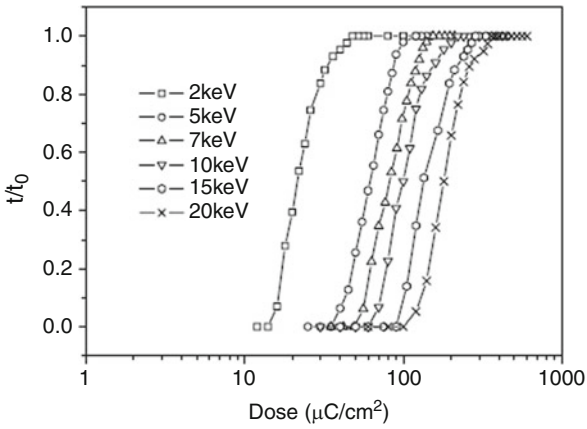


Fig. 3.28 Experimental development curves of HSQ negative tone resist exposed at different e-beam energies

Fig. 2.9 of Chap. 2). Similar to photoresists, long molecule chains are broken by energized electrons into short chains, called chain secession process in a positive tone resist, so that the exposed parts of polymer become soluble in developer. For a negative tone resist, the initial short chain molecules are joined upon exposure to form long chains, called cross-linking process, so that the exposure parts of polymer become insoluble in developer. The development curves, also called contrast curves, are obtained by exposure and development of a series of large square patterns with different exposure doses. The remaining resist thickness is measured against the exposure dose, which generates the curves shown in Fig. 3.27. For positive resist, there is an exposure dose at which the resist layer is fully developed to bottom. This exposure dose is defined as the resist sensitivity. The smaller this threshold dose, the higher the resist sensitivity. In Fig. 3.27a, resist A is more sensitive than resist B. For negative resist, the development curve looks like the plot C in Fig. 3.27b. With the increase of exposure dose, the degree of cross-linking increases, and more resist thickness remains after development, until the whole resist layer is completely cross-linked. The sensitivity of negative resist is defined by the exposure dose which results in more than half the resist thickness remaining due to cross-linking. A practical example of development curves is shown in Fig. 3.28, where the depth of resist development at different exposure dose is plotted for a negative tone resist, HSQ [44]. The development curves at different beam energy are also shown.

One must be aware that the above definition of resist sensitivity is only for comparison purpose. It is defined by exposure of much larger pattern area, which is rarely representative in actual e-beam lithography. The actual exposure dose needed for a particular pattern exposure can be quite different. It depends on the following factors:

- Electron energy. Higher electron energy requires higher exposure dose, resulting in low resist sensitivity, as is shown in Fig. 3.28.
- Resist thickness. The thinner the resist, the lower the resist sensitivity, because fewer electrons deposit their energy within the thin resist layer.

Table 3.2 Sensitivity and resolution of some e-beam resists

	Resist tone	Sensitivity[a] (μC/cm^2)	Resolution (nm)
PMMA	+	100	10
ZEP-520	+	30	10
ma-N 2400	–	60	80
EBR-9	+	10	200
PBS	+	1	250
COP	–	0.3	1000

[a]Sensitivity measured at 20 keV beam energy

- Pattern density. Exposure of dense patterns requires less exposure dose, because electron scattering from neighboring exposure areas contributes to the overall energy deposition. Similarly, isolated patterns require high exposure dose, because a portion of electrons have scattered outside the pattern area.
- Substrate material. High density substrate materials result in more backscattering electrons into resist layer, contributing to the resist exposure. Therefore, resist coated on a high density material appears more sensitive than on a low density material.
- Process conditions. The post-exposure baking conditions, in the case of chemically amplified resists, the strength of developer, and the temperature of developer can all have influence on the resist sensitivity, which will be discussed later.

For high resolution e-beam lithography, low sensitivity resists are preferred and also inevitable. Table 3.2 shows some widely used e-beam resists, which indicates the correlation between sensitivity and resolution capability. PMMA, the positive tone resist discovered more than 50 years ago, is still being used today although it is of very low sensitivity. PMMA is by far the only positive e-beam resist which has the highest resolution capability. The smallest feature size produced by e-beam lithography in PMMA is 4–5 nm, which was demonstrated more than 20 years ago [45, 46]. It must be noted that PMMA resist comes with different molecular weight. Typical commercial PMMA resists are of 950 K and 495 K molecular weights [47]. High molecular weight is generally of low sensitivity but high resolution capability. Another positive tone e-beam resist alternative to PMMA is the ZEP series (520A and 530A) developed by ZEON of Japan [48], which are of similar resolution capability but higher sensitivity than PMMA. Only one third of exposure dose is needed for ZEP resists, as shown in Table 3.2. They also have better etch resistance than PMMA, which is a key property for resist pattern transferring to substrate.

The best negative tone e-beam resist is hydrogen silsesquioxane (HSQ). HSQ is a liquid form of silicon dioxide initially developed at Dow Corning with the product name of flowable oxide (Fx-12). HSQ was first reported for e-beam lithography in 1998 [49]. Nowadays sub-10 nm resist line features can be achieved routinely with HSQ. There was also a report of 5–6 nm isolated line features produced in 30-nm-thick HSQ by e-beam lithography at 100 keV [9], as well as sub-10 nm dense line features in ultrathin (10 nm) HSQ layer [50, 51]. Although PMMA has similar or

slightly better resolution capability than HSQ, the line edge roughness (LER) in PMMA at sub-10 nm dimension is very high, whereas the LER in HSQ at sub-10 nm resolution is much lower [52]. The etch resistance of HSQ is also much higher than that of PMMA because of its inorganic nature. In these aspects, HSQ is a better e-beam resist for extremely high resolution e-beam lithography. Due to the success in e-beam lithography, HSQ is nowadays supplied as a commercial e-beam resist from a number of vendors worldwide. One should be aware that commercial HSQ has a storage lifetime of 6 months, beyond which the resist property may change. E-beam lithography experiments on aged HSQ revealed an increase of sensitivity, although the resolution capability was not notably compromised [53]. A dry powder form of HSQ (AQM-HSQ [54]) was developed, which has infinite long shelf life but same properties as the Dow Corning HSQ [55]. It should be mentioned that HSQ is not only widely used as e-beam resists but also being used for EUV optical lithography [56].

If lithography throughput is not a critical issue, particularly in R&D where generating nanoscale patterns is paramount, then there are some inorganic resist materials which are of very low sensitivity but very high resolution capability. Use of inorganic materials as resist in e-beam lithography dated back more than 40 years ago, when 8 nm features were made for the first time by e-beam exposure of metal thin film to form a carbon contamination pattern which acted as an ion sputter etching mask [57]. The exposure dose required to generate the carbon contamination feature was, however, as high as $10^7 \mu C/cm^2$. A number of other inorganic materials have been tried for e-beam lithography over the years, including metal halides (LiF, AlF_3, CaF_2, MgF_2, NaCl) with threshold exposure dose $>10^6 \mu C/cm^2$ and metal oxides (AlOx, ZnO, TiOx) with threshold exposure dose $>10^3 \mu C/cm^2$ [52]. Even ice has been tried as an e-beam resist [58]. The process of ice lithography is rather simple: introducing water vapor into an e-beam tool equipped with a cryostage to form a layer of ice on the substrate surface and exposing with e-beam to evaporate the ice selectively to form an ice pattern. The ice pattern can be further transferred to substrate by a lift-off process which will be introduced in Chap. 8. Rising the chamber temperature to room temperature, the remaining ice evaporates and the fabrication is completed. Although very high exposure dose is needed, for example, to evaporate 80-nm-thick ice, it needs $10^6 \mu C/cm^2$ exposure dose at 20 keV [59], the advantages of ice lithography are also remarkable. These include in situ patterning without taking samples out of vacuum chamber, no need for solvent development and patterning on nonflat substrate, etc. [60]. By repeating the freezing-evaporation cycle, 3D nanostructures can be built [61]. Ice lithography can be done not only with water vapor but also with organic vapor as long as they can freeze [62]. Lower exposure dose is needed for organic ice than for water ice [63]. There were also attempts to mix organic resists with inorganic materials to create quasi-inorganic resists, for example, fullerene molecules (C_{60}) mixed with ZEP-520 [64], PMMA [65], or NEB-22 [66] and silica nanoparticles mixed with ZEP-520 [67], which have improved the resolution capability and etch resistance of the resists, but at the cost of reduced resist sensitivity.

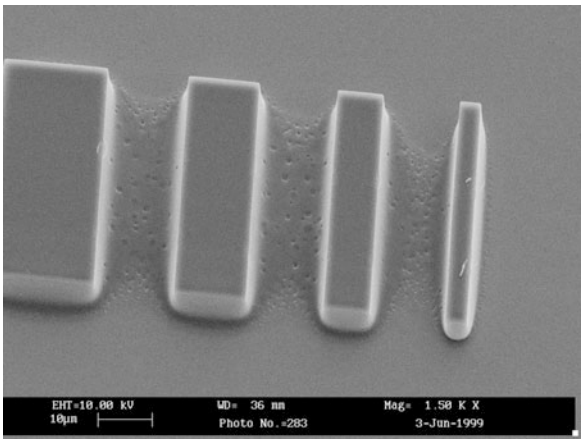


Fig. 3.29 SU-8 resist patterns exposed by e-beam at 50 keV energy

To understand why low sensitivity resists have high resolution capability, one should look at the PSF shown in Fig. 3.23. The long tail of PSF caused by secondary electrons and backscattered electrons will not have significant exposure effect if the resist is less sensitive. In other words, these low energy electrons will only show up their impact on resist pattern profile if the resist sensitivity is very high. An example is shown in Fig. 3.29 which is the SEM image of exposed SU-8 resist [68]. SU-8 is a photoresist but also sensitive to electrons. The sensitivity of SU-8 is around $1\mu C/cm^2$ at 50 keV beam energy, which is extremely high compared with PMMA. The consequence of such a high sensitivity is the residues of SU-8 resist shown between the exposed resist patterns. The residues are caused by exposure of backscattering electrons. They only exist between the exposure patterns because of the accumulated effect from both sides of electron scattering. If the resist had not been so sensitive, the residues would not have occurred.

Normally a pattern is exposed by e-beam lithography through multi-pass scan to fill up the exposure area, because the beam diameter and scan grid are much smaller than the pattern dimension. For very high resolution lithography of line features ($<$50 nm), a single pass exposure is often employed, whereby the electron beam scans only a single pass. The linewidth is then determined by the exposure dose of this single pass. Although a single pass exposure should theoretically be only the width of beam diameter, the actual line in resist can be much wider if exposed at high dose.

Because the accuracy of resist pattern is so much dependent on exposure dose, it is a common practice that e-beam lithography of any new pattern designs is always started with a dose test, i.e., running a series of exposure for the same pattern design at different exposure doses, from which the optimum exposure dose range is determined.

3.5.2 Contrast

Resist contrast is defined by the slope of a development curve and expressed in Eq. 3.17. D_0 is the exposure dose at which the resist starts to be exposed, also sometimes called threshold dose (D_{th}), and D_1 is the exposure dose at which the resist is fully exposed. They are marked in Fig. 3.27a. The steeper the slope is, the higher the resist contrast. For example, the resist A in Fig. 3.27b has higher contrast than resist B. Resist contrast is an inherent property of a resist material. However, development conditions can have some influence on it.

$$\gamma = \frac{1}{\log_{10} D_1 - \log_{10} D_0} \tag{3.17}$$

By this definition, the contrast of negative tone resist HSQ shown in Fig. 3.28 is 2.8 [44]. Another negative tone resist, SU-8, has a contrast of only 0.92, as measured from contrast curve shown in Fig. 3.30 [68, 69]. By the appearance of contrast curves of these two resists, one can immediately tell that the HSQ resist has higher contrast. Apart from measuring the slope of resist contrast curves, another direct way of judging the resist contrast is to observe the resist profile after development. An example to compare the effect of resist contrast on resist profile is shown in Fig. 3.31, where computer-simulated resist profiles are shown for 50 keV e-beam exposure of 0.5-μm-thick resist. Figure 3.31a is the profile of low contrast resist, while Fig. 3.31b is the profile of high contrast resist. Although electron scattering and proximity effect are the same in both cases, high contrast resist can produce the most desirable profile. High contrast enables resist profiles to have vertical sidewalls and high aspect ratios (feature height to feature width), while low contrast resist will have tapered sidewall. As an example, e-beam exposure of low contrast SU-8 resist with sensitivity of 1μC/cm^2 resulted in sidewall broadening which is proportional to

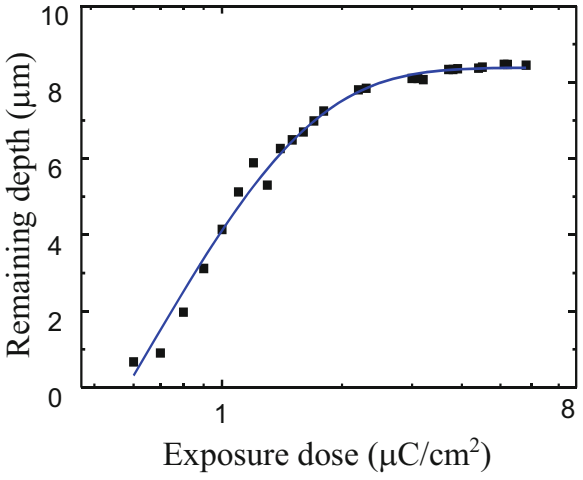


Fig. 3.30 Contrast curve of SU-8 resist

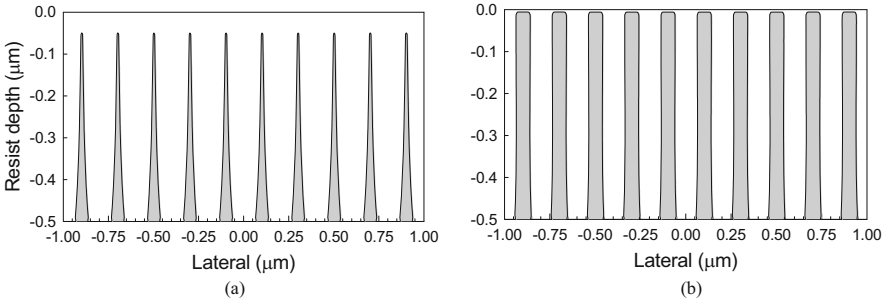


Fig. 3.31 (**a**) Low contrast resist profile, (**b**) high contrast resist profile

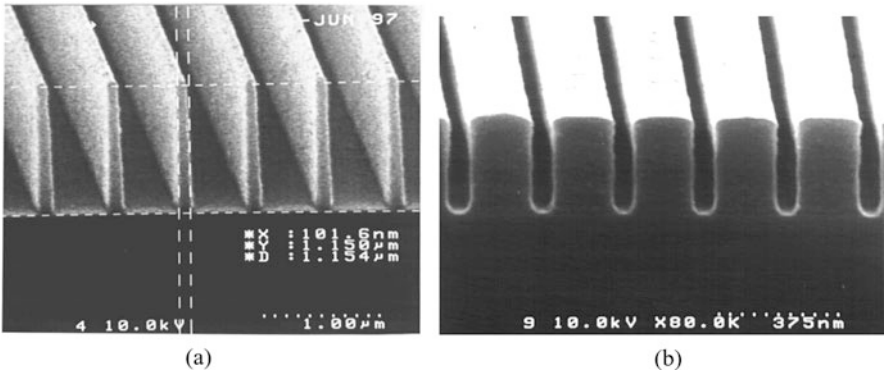


Fig. 3.32 (**a**) AZPN114 resist profiles (100 nm lines with aspect ratio 11:1), (**b**) UV3 resist profiles (40 nm lines with 8:1 aspect ratio)

the increase of exposure dose. The broadening at the bottom of 1 μm line structure was as large as 2.7 μm when exposed at 6μC/cm^2 and became 5.5 μm at 18μC/cm^2 [70].

As previously mentioned, generally low sensitivity resists have high contrast and high sensitivity resists have low contrast. However, there is an exception that some chemically amplified (CA) resists are of both high sensitivity and high contrast. CA resists have been introduced in Sect. 2.3.2 in Chap. 2. They are highly sensitive because a few photoacid generators (PAGs) can be amplified in numbers at later stage of post-exposure baking. The CA resists were originally developed for deep UV optical lithography but later were found also sensitive to electrons, hence used as e-beam resists as well. Figure 3.32 shows high aspect ratio patterns in negative tone CA resist AZPN114 (Fig. 3.32a) [71] and positive tone CA resist UV3 (Fig. 3.32b) [72] exposed by 100 keV electron beam. Similar results have been demonstrated for other CA resists, such as SAL601 [71] and NEB-22 [73]. These CA resists are generally capable of resolution below 50 nm. Although their resolution capability is not as high as PMMA or HSQ, their very high sensitivity, two to three orders of

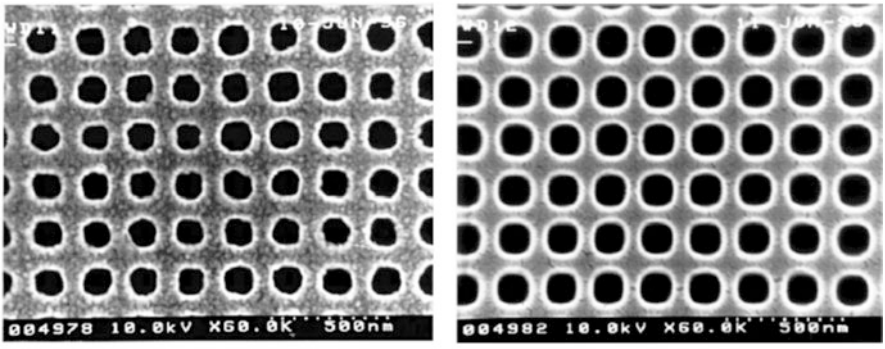


Fig. 3.33 100 nm holes exposed with (**a**) PMMA at 1450μC/cm^2 and (**b**) AZPF514 at 25μC/cm^2

magnitude higher, makes them very attractive in high-throughput e-beam lithography. For example, Fig. 3.33 shows the 100 nm hole array exposed by 100 keV electron beam with PMMA at the exposure dose of 1450μC/cm^2 (Fig. 3.33a) and with AZPF514, a positive tone CA resist, at the exposure dose of 25μC/cm^2 (Fig. 3.33b) [72]. The CA resist is nearly 60 times more sensitive than PMMA.

In the early days when optical lithography stopped at 193 nm wavelength, the IC industry was seriously looking at the possibility of using high-throughput e-beam lithography for further technology nodes. Unfortunately, optical lithography marched on with immersion and multi-patterning technologies and e-beam direct write for mass production of ICs never took off. On the one hand, the CA resists have many issues such as highly sensitive to post-exposure baking condition, post-exposure delay effect, and susceptible to exposure environment [71, 74, 75]. On the other hand, scientific community only cares about high resolution, not productivity. Researchers worldwide mostly contend with a few mature e-beam resists such as PMMA and HSQ for nanoscale patterning, though there are still people interested in research on new CA resists purely from scientific point of views [76–78]. One must be aware that resists of too high sensitivity may not be suitable for some slow scanning e-beam systems. If the scanning speed of an e-beam system is slow, for example, below 10 MHz, the beam current has to be extremely small in order not to overexpose some very high sensitivity CA resists. Extremely small beam current makes the focus calibration and registration extremely difficult because of low secondary electron image. Some systems may have a lower limit on attainable beam current, which renders the exposure of CA resists by these systems impossible.

3.5.3 Resolution Enhancement Processes

The resist sensitivity and contrast are partly dependent on the resist material property, as well as partly dependent on the resist process conditions. Although every

commercial resist comes with a guideline process conditions given by resist vendors, the conditions are meant only for general applications and may not be tuned to achieve the highest resolution capability. The process conditions are also tied to pattern density and exposure condition, such as the beam energy and exposure dose. Therefore, to achieve the highest resolution for a particular resist, the optimum process conditions should be investigated. In some cases, exotic methods were developed to enable very high resolution of e-beam lithography which has previously been unthinkable. These exotic methods include ultrasonic assisted resist development and development at lower or higher temperature.

The enhancement effect by ultrasonic assisted development on e-beam lithography resolution with PMMA resist was discovered in 1993 [79]. Previously it had been generally accepted that e-beam lithography has a resolution limit at around 10 nm because of secondary electron scattering effect [80]. However, one fact was overlooked in considering the resolution limit, that is, the behavior of polymer molecules during the development process. The development is realized by short chain polymer molecules dissolved in a developer solution after chain scission action by electron exposure. When the exposure region becomes very small (<10 nm), the molecules, which for PMMA is of a size around 2 nm, can be trapped in solid PMMA due to intermolecular forces which increase rapidly within the radius of 10 nm. More electron interaction with the polymer is needed to release the trapped polymer molecules, which means higher exposure dose. The exposure region then becomes wider at the increased exposure dose. With ultrasonic assisted development, the ultrasonic agitation can provide additional energy and help to loosen up the trapped molecules to dissolve into the developer solution. With the assistance of ultrasonic agitation, exposure lines as small as 4 nm in PMMA has been achieved [45]. The ultrasonic agitation was also proved to reduce the exposure dose significantly at feature size below 20 nm, as shown in Fig. 3.34. It becomes no difference to conventional development process for big feature size because the intermolecular

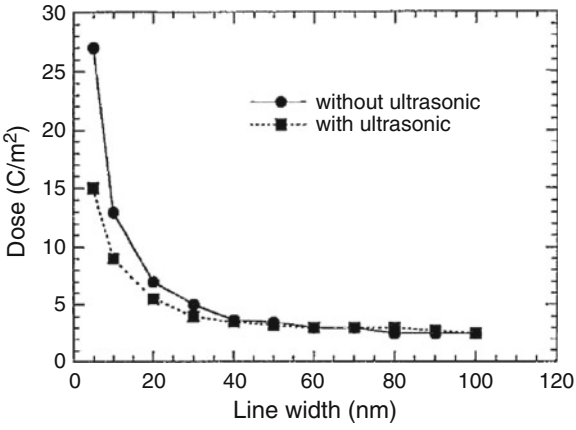


Fig. 3.34 Comparison of exposure dose required for PMMA development with and without ultrasonic assistance. (Reprint from [79] with permission)

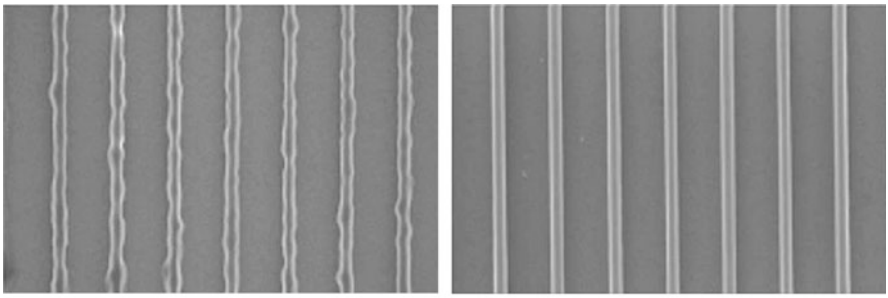


Fig. 3.35 Comparison of edge roughness of ZEP-520 resist lines (40 nm wide) developed at (**a**) room temperature and (**b**) at −4 °C temperature. (Reprint from [82] with permission)

force is no longer dominant at above 10 nm line opening. There were also other explanations as to why the ultrasonic agitation can help enhance the resolution of PMMA, such as reduction of developer viscosity or reduction of resist swelling [81]. However, it must be pointed out that the resolution enhancement by ultrasonic agitation only works for PMMA resist with high molecular weight, which demonstrates that resist development process is very much dependent on individual resist materials.

Another method to enhance resolution is to develop the exposed PMMA in a cold developer. Sub-10 nm lines were achieved at 30 keV beam energy with MIBK+IPA (1:3) developer which was maintained at 4–8 °C temperature [46]. Unlike ultrasonic agitation, which promotes the activity of PMMA molecules, cold developer actually suppresses the activity of PMMA molecules. In order to develop the exposed PMMA resist at low temperature, exposure dose much higher than development in room temperature is needed. High exposure dose is known to result in wider exposure region, as indicated by previous electron scattering analysis. However, energy level deposited in resist is rapidly reduced at the location away from the center of exposure region, as shown by the point spread function for very thin resist layer in Fig. 3.13. In a cold developer, those PMMA molecules exposed by low level electron energy cannot be dissolved because of their low level of kinetic activity at low temperature. Only the very central region where the deposited electron energy is the highest can be developed, resulting in high resolution exposure features. The similar resolution enhancement at low temperature development was also found in e-beam lithography of ZEP-520 resist. With the developer temperature as low as −17 °C, 13 nm lines in ZEP-520 exposed at 30 keV beam energy have been obtained [82]. A particular benefit of low temperature development for ZEP-520 is the much improved line edge roughness, as demonstrated in Fig. 3.35. The ZEP-520 resist lines in Fig. 3.35a were developed at room temperature, while the lines in Fig. 3.35b were developed at −4 °C. It clearly shows the ZEP-520 resist molecules have much lower kinetics in low temperature developer solution.

The reason that cold development works for both PMMA and ZEP-520 resists is believed to be due to the variation of molecule weight distribution before and after

Table 3.3 Influence of developer concentration on the resolution of PMMA resist

Developer concentration(MIBK-IPA)	Sensitivity	Resolution
1:3	Low	Extremely high
1:2	Medium	Very high
1:1	High	High
Pure MIBK	Very high	Low

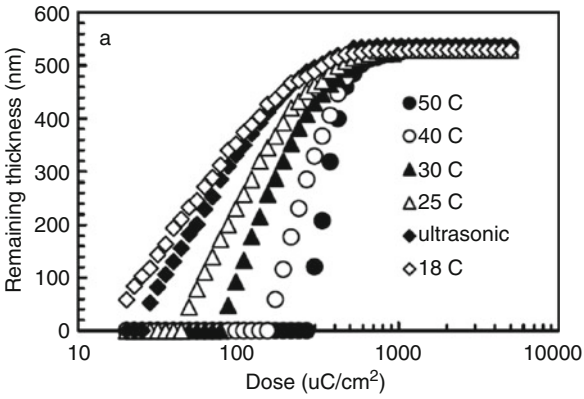


Fig. 3.36 Contrast curves of HSQ resist at different development temperatures

electron exposure. Upon electron exposure, chain scissions occur, resulting in short chains and low-weight molecules in the exposed region. Low temperature developer reduces the kinetic energy of molecules. Only those of molecules with very short chain can be dissolved by the developer, which effectively suppressed the influence of scattering electrons because the polymer molecules exposed by low energy scattering electrons cannot participate the dissolving process. The same mechanism can explain why low strength developer can help to achieve high resolution resist patterns. Table 3.3 shows qualitatively the relation between developer concentration and resolution of PMMA resist. On the other hand, positive chemically amplified resists such as UV3 do not show any improvement in low temperature developers because there is no molecular weight degradation process during e-beam exposure of these resists [82]. However, cold development does show better results for some negative chemically amplified resists such as NEB-22 [83] and SAL-601 [84].

While low temperature development helps to enhance resolution of PMMA and ZEP-520, it works the opposite for HSQ resist. The contrast of HSQ resist is actually enhanced at high development temperature, as shown in Fig. 3.36, where development temperature varies from 18 °C to 50 °C, including ultrasonic assisted development [85]. The sensitivity of HSQ is decreasing but the contrast is increasing with the increase of development temperature. It is not clear exactly why the high temperature development has caused the decrease of e-beam exposure sensitivity of HSQ, as most other resists become more sensitive at high development

temperature. However, the benefit of high temperature development of HSQ is apparent. High resolution dense patterns can be resolved when developed at high temperature. However, there is a limit on how high the development temperature should be. It is found there is a significant destruction of the exposed structures in the resist at development temperature of 60 °C [86].

3.6 Ultimate Resolution of E-Beam Lithography

E-beam lithography is by far of the highest resolution, most versatile, and most widely used lithography technique for nanoscale patterning. No wonder that every laboratory around the world which claims to be a nanofabrication lab will be found to have an e-beam lithography system, whether it is a multimillion dollar complex machine or a simple SEM converted system. There are many factors, in addition to the hardware capability, that influence the ultimate resolution an e-beam machine can achieve. These factors have been analyzed in previous sections. Here a summary of conditions or preferences is given if one intends to achieve the highest possible resolution of e-beam lithography.

1. Using high electron energy. It can reduce the spread of forward scattering electrons and enlarge the spread of backscattering electrons. The latter can avoid local concentration of backscattering electron energy deposition and both can effectively reduce the proximity effect. At high beam energy, both the chromatic aberration and space charge effect in the electron lens system are reduced, which helps to obtain small beam spot. High energy electrons also penetrate deeper in a resist layer, which facilitate the exposure of thick resist and to achieve high aspect ratio resist pattern structures.
2. Choosing small aperture. The aperture that limits the angle of electron beam spread in an e-beam system is generally selectable. For high resolution e-beam lithography, small aperture is preferred, which can effectively reduce the spherical aberration since the aberration is proportional to the cubic power of beam spread angle.
3. Using low beam current. Low beam current can reduce the aberration caused by space charge effect and helps to reduce the beam spot size. However, the adverse effect of low beam current is the increase of exposure time as well as reduced electron signal intensity which is used to detect the focusing and overlay marks.
4. Choosing small scanning field. Small scanning field means small-angle deflection of beam, which can reduce the aberrations associated with deflection.
5. Using low sensitivity e-beam resists. Low sensitivity resists can effectively suppress the exposure of resist by secondary electrons and low energy backscattered electrons. For lithography below 20 nm, low sensitivity resists such as PMMA (positive tone) and HSQ (negative tone) are commonly used.

6. Using thin layer resist. Thin resist can reduce the proximity effect caused by forward scattering electrons.
7. Optimizing resist processes. If the above conditions cannot all be met, optimized resist processes including resist coating, adhesion, development condition, and pre-/post-baking can to some extent help to improve the lithography resolution.
8. Avoiding high density patterns. As previously explained, the most difficult pattern to achieve high resolution is the high density pattern. A single 10 nm line pattern is far easier than dense pattern of 10 nm half pitch to obtain. Even proximity effect correction is powerless for high resolution periodic dense patterns.
9. Avoiding high density or poor conducting substrates. High density substrates generate more backscattering electrons, which means more severe proximity effect; poor conducting substrate causes accumulation of electrical charge which generates local electrical field, causing distortion of exposed patterns, as well as excessive heat if the beam current is high.
10. The last but not the least is to maintain a stable environment for the e-beam system to operate. The stable environment includes low surrounding vibration, low electromagnetic interference, and low fluctuation of ambient temperature. These conditions become necessary when exposing sub-100 nm patterns.

What is the ultimate resolution of e-beam lithography? Such a question has been asked over and over, and many researchers have been trying to demonstrate the smallest feature one can achieve [5, 34, 80]. Clearly, the beam size is not a limiting factor, as the electron optical system in a scanning transmission electron microscope (STEM) can achieve the beam size well below 1 nm. An experiment has demonstrated 1.7 nm line feature in PMMA using an aberration-corrected STEM [87]. In the experiment, the electron energy was 200 keV which enabled 0.5 nm beam size; the substrate was a 5-nm-thick SiNx membrane which minimized the electron backscattering effect; the PMMA is 7 nm thick which minimized the forward scattering effect. Another experiment using aberration-corrected STEM at 200 keV and beam size of 1.5 nm demonstrated 2 nm feature size in HSQ which was 10 nm thick on a 10-nm-thick SiNx membrane [88]. Theoretically the ultimate resolution should only be limited by beam diameter [5]. In practice, many factors can influence the resolution, such as the aforementioned ten factors. One should bear in mind that the goal of electron beam lithography is for nanofabrication. In most cases, the patterned resist features have to be transferred to substrate either through etching or lift-off [89], which limits the ultimate resolution of patterning common e-beam resists PMMA or HSQ by e-beam lithography around 4–5 nm, though there are indirect techniques which can pattern even smaller features as will be introduced in Chap. 9.

3.7 High-Throughput E-Beam Lithography

Although an electron beam can be focused into a very small spot and, therefore, can do very high resolution lithography, its fundamental limit is the low throughput because of its sequential nature in exposure. While optical lithography has standard throughput of at least 60 wafers per hour, a single 50 nm electron beam has to have a scanning speed of 20 km/s and beam dwelling time of 2.5 ps in order to compete with this throughput [90]. None of the single electron beam lithography systems can ever dream of such a high scanning speed, and none of the e-beam resists have such high sensitivity. At the current level of scanning speed and beam current for a Gaussian beam system, an 8 inch wafer would need tens of hours to expose. To make a photomask for 22 nm node of IC patterning with 193 nm photolithography, the e-beam writing time would be 80 hours [91].

The key motivation to develop high-throughput e-beam lithography systems in the early days was to prepare for the next generation lithography (NGL) after optical lithography ran out of steam for 16 nm generation or even single digit feature size of ICs. By 2010, the semiconductor industry considered that 193 nm optical lithography with immersion imaging and multi-patterning could go as far as 22 nm node of IC patterning. Below 22 nm, EUV was not ready to take over and e-beam direct write might have the chance to play a role in mass production of ICs at 16 nm or even 11 nm [92]. A single beam system can never meet the challenge. Electron beam projection lithography became the favorable option. While optical lithography relies on mask projection, e-beam can do the same. In e-beam mask projection lithography, a broad beam is used as the "light" source to illuminate a mask. Electrons passing through the mask are focused into high resolution image to expose a resist layer. Several types of e-beam projection lithography (EPL) systems were developed including the cell projection system from Hitachi (HL-800D) [93], the projection reduction exposure with variable axis immersion lens (PREVAIL) system from IBM [94], and the scattering with angular limitation projection electron-beam lithography (SCALPEL) system from Lucent Technologies [95]. All of these e-beam lithography technologies rely on exposure through a mask. However, the mask, whether stencil type or membrane type, is difficult to make, to inspect, and to repair. This has been an age-old problem when X-ray lithography was in its peak development era. Mask projection lithography has failed to be adopted in industrial manufacturing of ICs.

The situation has changed fundamentally since 2016. While 193 nm immersion lithography combined with multi-patterning continued to extend ICs to 10 nm node, EUV technology became mature and started IC production in 2019, taking over from 7 nm node onward [96]. High-throughput e-beam lithography is now only for making masks. There are two types of e-beam lithography technologies active in advanced mask making: variable shaped beam (VSB) and multi-beam mask writer (MBMW).

3.7.1 Variable Shaped Beam Lithography

The high-throughput e-beam lithography started with shaped beam system in the 1970s [97]. The so-called shaped beam is to use a shaped aperture, usually a square shape so that the actual pixel exposed onto a wafer is no longer a single Gaussian beam of very fine size but a large beam. As mentioned in Sect. 3.3.4, a pattern is always fragmented into basic shapes of rectangles, trapeziums, or parallelograms in e-beam lithography. Instead of scanning a Gaussian beam to fill the shapes, the shaped beam can fill them much quick, which considerably increases the exposure speed, hence the throughput. To enable the flexibility of exposing various pattern size, variable shaped beam (VSB) was developed [98]. VSB is an extension of the fixed shape beam concept. As illustrated in Fig. 3.37, the beam spot is shaped by projecting two square apertures simultaneously. By shifting the image of the first aperture laterally with respect to the second aperture, a compound beam spot is formed, which is subsequently demagnified and projected onto the target. By the combination of two square apertures, the beam spot is tailored to fit each pattern segment while it steps from one segment to the next. A typical VSB system operates at 50 keV beam energy and basic beam size of 2 × 2 μm. Originally VSB was developed mainly to prepare for high-throughput patterning of ICs. Lately VSB became the powerful tool for mask making as laser direct write could not meet the mask requirements below 130 nm IC technology. Currently, there are three major

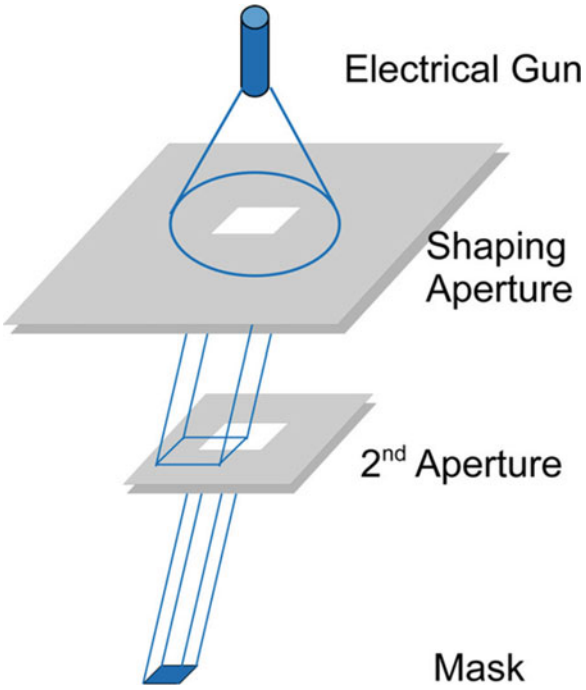


Fig. 3.37 Combination of apertures for variable shaped beam lithography

types of VSB systems worldwide, i.e., JBX series from JEOL [99], SB series from Vistec [100], and EBM series from NuFlare [101]. These VSBs aim at masks for 22 nm node down to 7 nm node, noted that mask feature is four times of IC feature.

3.7.2 Multi-E-Beam Lithography

The advance in optical lithography (193 nm immersion + multi-patterning and EUV) has left the high-throughput e-beam lithography with the only option: mask writing. While e-beam mask writers were traditionally monopolized by VSB technology as described in previous section, the ever-increased complexity of 193 nm and EUV lithography add more and more sub-resolution features onto the masks; in particular, the adoption of inverse lithography technology (ILT) prefers curvilinear instead of rectangular features [102]. To pattern the curvilinear features, the size of shaped beam in a VSB has to be extremely small, which leads to considerable increase of shots (exposure points), hence the exposure time. It is estimated that the e-beam writing time for an ILT mask would be more than 30 hours, whereas the writing time could be reduced to less than 10 hours by using a multi-beam mask writer [103]. Three different types of multi-e-beam lithography systems were being developed in the past, i.e., projection maskless lithography (PML2) [104], reflective e-beam lithography (REBL) [105], and high-throughput maskless lithography (MAPPER) [106]. Fast-forward to present day, the company which developed the MAPPER system was bankrupt in 2018 and taken over by ASML in 2019. KLA-Tencor which was behind the REBL is no longer supporting its development. Only the PML2 was continuously developed by IMS (IMS Nanofabrication AG, Austria) which gave up on e-beam direct write for IC patterning but teamed up with JEOL to focus on mask making. The system later became MBMW-101, world's first high-throughput multi-beam mask writer [107].

The MBMW system uses array of microscale beam blanking apertures, called programmable Aperture Plate System (APS), as schematically shown in Fig. 3.38 [108]. The APS is the combination of an aperture plate and a blanking plate with IC control electronics. The aperture plate splits a broad telecentric 5 keV electron beam into thousands of 2.5-μm-sized beamlets. These individual beamlets then pass through the blanking plate where each opening aperture is integrated with a deflection electrode controlled by the IC electronics. Unwanted beamlets are slightly deflected and filtered out at the second cross-over of the projection optics. Therefore only beamlets which are unaffected by the blanking plate are projected to the wafer substrate with 200 times reduction to achieve 12.5 nm spot size and accelerated to 50 keV beam energy [109]. Unlike the VSB which only has a single beam, MBMW has the total number of beams from 256 k to one million, depending on the number of apertures, which allow it to pattern millions of features simultaneously and therefore can considerably increase the throughput.

Another multi-beam technology (MBW series) is developed by NuFlare Technology, as a direct competition to MBMW-101. The MBW system is based on the

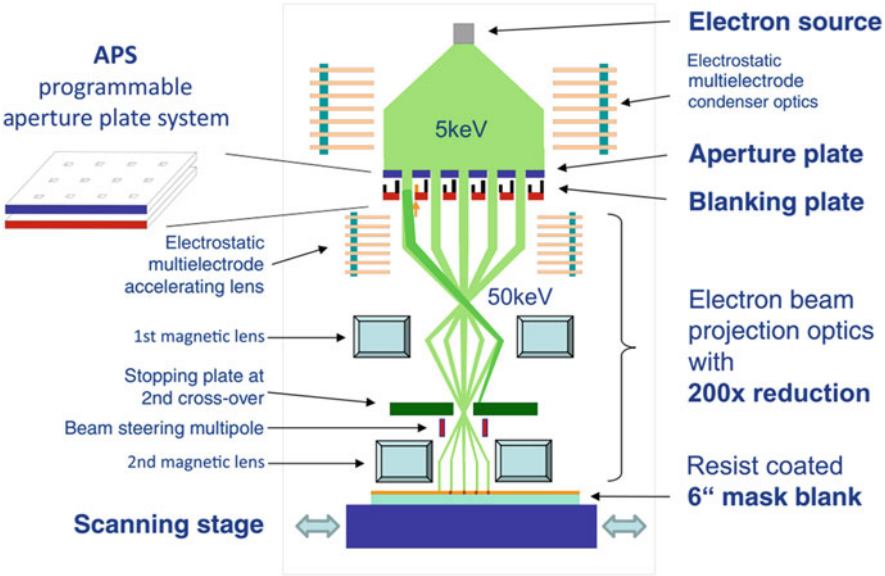


Fig. 3.38 Schematics of multi-beam mask writer developed by IMS [108]

same concept of dividing a broad beam into multiple individual beams (260,000) by a shaping aperture array (SAA) and deflecting individual beam by a blanking aperture array (BAA) [110]. The only difference is that the MBW system deflects 50 keV beams, while the MBMW system deflects 5 keV beams. Direct blanking of a high energy beam enables the system to be less sensitive to external noise in all beam paths. It also enables simple and robust electrical insulation design because it does not need an additional high voltage area that may cause unexpected discharges.

With wide adoption of ILT from 5 nm node, multi-e-beam technology becomes the only option enabling both EUV and curvilinear photomasks. The eBeam Initiative's 11th annual Luminaries survey in 2022 reported strong purchasing predictions for multi-beam mask writers, and the multi-beam mask writing technology is hailed as "a game changer" [111].

References

1. Haller, I., Hazakis, M., Srinivasan, R.: High resolution positive resists for electron beam exposure. IBM J. Res. Develop. **12**, 251 (1968)
2. Grivet, P.: Electron Optics. Pergamon Press, New York (1965)
3. Mayadas, A.F., Laibowitz, R.B.: One-dimensional superconductors. Phys. Rev. Lett. **28**, 156 (1972)
4. Hutcheson, G.D.: E-beam direct write: why it's always left standing at the altar of new nodes. Proc. SPIE. **10584**, 2018 (2018)

5. Jiang, N.: On the spatial resolution limit of direct-write electron beam lithography. Microelectron. Eng. **168**, 41–44 (2017)
6. Harris, L.A.: Introduction to Electron Beam Technology, R. Bakish (eds.). John Wiley & Sons, New York (1962)
7. Cui, Z., Tong, L.: Optimum geometry and space charge effect in vacuum microelectronic devices. IEEE Electron Devices. **ED-40**(2), 448 (1993)
8. GEMINI field emission electron optics: core technology for the LEO SUPRA range. Available from: www.smt.zeiss.com
9. Yamazaki, K., Namatsu, H.: 5-nm-order electron-beam lithography for nanodevice fabrication. Jpn. J. Appl. Phys. **43**(6B), 3767 (2004)
10. Hawkes, P.W., Kasper, E.: Principles of electron optics. Academic Press, London (1996)
11. Rouse, J., Zhu, X., Munro, E.: Solution of electron optics problems with space charge in 2D and 3D. Proc. SPIE. **2522**, 375–386 (1995)
12. Broers, A.N.: High-resolution thermionic cathode scanning transmission electron microscope. Appl. Phys. Lett. **22**, 610–612 (1973)
13. Bracher, B.H., Prewett, P.D.: Electron beam spot size calculation and software. In: Rutherford Appleton Laboratory Internal Report (1997)
14. Munro, E.: Electron and ion optical design software for integrated circuit manufacturing equipment. J. Vac. Sci. Technol. **B15**(6), 2692 (1997)
15. Hara, Y.: Electron beam projection litho system poised for test. EE Times, July 2003
16. Khodadad, I., et al: Electron beam lithography using fixed beam moving stage. J. Vac. Sci. Technol. **B**35(5), 051601 (2017)
17. Yallew, H.D., Jágerská, J., Greve, M.M.: Long, stitch-free slot waveguide with s-bend tapered couplers for IR-sensing applications using electron beam lithography. J. Vac. Sci. Technol. B. **41**, 012601 (2023)
18. EBPG Plus: High-resolution lithography with automation, throughput, and reliability. https://raith.com/
19. JBX-9500FS electron beam direct write system. www.jeol.co.jp (2015)
20. Variable shaped beam (VSB) technology from Vistec. https://www.vistec-semi.com/
21. NanoBeam nB5. http://www.nanobeam.co.uk/ (2015)
22. Nanopatterning made simple by ELPHY upgrade kits. https://raith.com/product/elphy/
23. McCord, M.A., Rooks, M.J.: Electron beam lithography. In: Rai-Coudhury, P. (ed.) Handbook of Microlithography, Micromachining and Microfabrication. SPIE Optical Engineering Press (1997)
24. Hawryluk, R.J., Hawryluk, A.M., Smith, H.I.: Energy dissipation in a thin polymer film by electron beam scattering. J. Appl. Phys. **45**(6), 2551 (1974)
25. Murata, K., Matsukaw, T., Shimizu, R.: Monte Carlo calculations on electron scattering in a solid target. Jpn. J. Appl. Phys. **10**(6), 678 (1971)
26. Parikh, M., Kyser, D.F.: Energy deposition function in electron resist films on substrates. J. Appl. Phys. **50**(2), 1104 (1979)
27. Cui, Z.: MOCASEL: a total solution to electron beam lithography simulation. Proc. SPIE. **3676**, 494 (1999)
28. Parikh, M.: Self-consistent proximity effect correction technique for resist exposure (SPECTRE). J. Vac. Sci. Technol. **15**(3), 931 (1978)
29. Parikh, M.: Correction to proximity effects in electron beam lithography. II. Implementation. J. Appl. Phys. **50**(6), 4378 (1979)
30. Hintermaier, M., et al.: Proximity correction using computer aided proximity correction (CAPROX) evaluation and application. J. Vac. Sci. Technol. **B9**(6), 3043 (1991)
31. Owen, G., Rissman, P.: Proximity effect correction for electron beam lithography by equalization of background dose. J. Appl. Phys. **54**(6), 3573 (1983)
32. Jackel, L.D., et al.: Beam energy effects in electron beam lithography: the range and intensity of backscattered exposure. Appl. Phys. Lett. **45**, 698 (1984)

33. Wu, B., Neureuther, A.R.: Energy deposition and transfer in electron beam lithography. J. Vac. Sci. Technol. **B19**(6), 2508 (2001)
34. Kyser, D.F.: Spatial resolution limits in electron beam lithography. J. Vac. Sci. Technol. **B1**(4), 1391 (1983)
35. Xia, X., et al.: Fabrication of near-infrared and optical meta-materials on insulating substrates by Lift-off using PMMA/Al stack. Microelectron. Eng. **84**(5–8), 1144–1147 (2007)
36. Cui, Z.: TEMP: a software package for simulating resist heating in e-beam lithography. Proc. SPIE. **3331**, 420–430 (1998)
37. Babin, S., Hudek, P., Kosic, I.: Quantitative measurement of the resist heating in a variable shaped electron lithography. J. Vac. Sci. Technol. **B15**(2), 311 (1997)
38. Yang, H., et al.: Low-energy electron-beam lithography of ZEP-520 positive resist. In: The 1st IEEE International Conference on Nano/Micro Engineered and Molecular Systems, pp. 391–394. Zhuhai, China (2006)
39. Lee, Y.H., et al.: Low voltage alternative for electron beam lithography. J. Vac. Sci. Technol. **B10**(6), 3094 (1992)
40. Yoshizawa, M., et al.: Challenges to ultra-thin resist process for LEEPL. J. Photopolym. Sci. Technol. **17**(4), 581–586 (2004)
41. McCord, M.A., Newman, T.H.: Low voltage, high resolution studies of electron beam resist exposure and proximity effect. J. Vac. Sci. Technol. **B10**(6), 3083 (1992)
42. Utsumi, T.: Low-energy e-beam proximity lithography (LEEPL): is the simplest the best? Jpn. J. Appl. Phys. **38**(12B), 7046 (1999)
43. Utsumi, T.: Present status and future prospects of LEEPL. Microelectron. Eng. **83**, 738–748 (2006)
44. Yang, H., et al.: Low-energy electron-beam lithography of hydrogen silsesquioxane. Microelectron. Eng. **83**, 788–791 (2006)
45. Yasin, S., Hasko, D.G., Ahmed, H.: Fabrication of < 5 nm width lines in poly (methylmethacrylate) resist using a water: isopropyl alcohol developer and ultrasonically-assisted development. Appl. Phys. Lett. **78**(18), 2760–2762 (2001)
46. Hu, W., et al.: Sub-10 nm electron beam lithography using cold development of poly (methylmethacrylate). J. Vac. Sci. Technol. **B22**(4), 1711 (2004)
47. PMMA electron beam resists product data sheets. https://kayakuam.com/
48. ZEP series positive tone electron beam resists from ZEON. https://www.zeon.co.jp/
49. Namatsu, H., et al.: Nano-patterning of a hydrogen Silsesquioxane resist with reduced linewidth fluctuation. Microelectron. Eng. **41–42**, 331–334 (1998)
50. Grigorescu, A.E., et al.: 10 nm lines and spaces written in HSQ, using electron beam lithography. Microelectron. Eng. **84**(5–8), 822–824 (2007)
51. Grigorescu, A.E., et al.: Influence of the development process on ultimate resolution electron beam lithography, using ultrathin hydrogen silsesquioxane resist layers. J. Vac. Sci. Technol. B. **25**(6), 1998–2003 (2007)
52. Grigorescu, A.E., Hagen, C.W.: Resists for sub-20-nm electron beam lithography with a focus on HSQ: state of the art. Nanotechnology. **20**, 292001 (2009)
53. Chen, X., et al.: Aged hydrogen silsesquioxane for sub-10 nm line patterns. Microelectron. Eng. **163**, 105–109 (2016)
54. AQM-HSQ available from Applied Quantum Materials, Inc. https://www.aqmaterials.com/
55. Shen, J., et al.: E-beam lithography using dry powder resist of hydrogen silsesquioxane having long shelf life. J. Vac. Sci. Technol. B. **37**, 021601 (2019)
56. Rathore, A., et al.: Extreme ultraviolet-printability and mechanistic studies of engineered hydrogen Silsesquioxane photoresist systems. ACS Appl. Polym. Mater. **3**(4), 1964–1972 (2021)
57. Broers, A.N., et al.: Electron-beam fabrication of 8nm metal structures. Appl. Phys. Lett. **29**(9), 596–598 (1976)
58. King, G.M., et al.: Nanometer patterning with ice. Nano Lett. **6**, 1157–1160 (2005)
59. Han, A., et al.: Ice lithography for nanodevices. Nano Lett. **10**, 5056–5059 (2010)

60. Yao, G., et al.: Ice-assisted electron-beam lithography for MoS_2 transistors with extremely low-energy electrons. Nanoscale Adv. **4**, 2479 (2022)
61. Hong, Y., et al.: Three-dimensional in situ electron-beam lithography using water ice. Nano Lett. **18**, 5036–5041 (2018)
62. Tiddi, W., Elsukova, A., Beleggia, M.: Organic ice resists for 3D electron-beam processing: instrumentation and operation. Microelectron. Eng. **192**, 38–43 (2018)
63. Zhao, D., et al.: Ice lithography for 3D nanofabrication. Sci. Bull. **64**, 865–871 (2019)
64. Ishii, T., Nozawa, H., Tamamura, T.: Nanocomposite resist system. Appl. Phys. Lett. **70**(9), 1110 (1997)
65. Dentinger, P.M., Taylor, J.W.: Increasing plasma etch resistance of resists using fullerene additives. J. Vac. Sci. Technol. **B15**(6), 2575 (1997)
66. You, H.-C., Ko, F.-H., Lei, T.-F.: Resist nano-modification technology for enhancing the lithography and etching performance. Microelectron. Eng. **78–79**, 521–527 (2005)
67. Merhari, L., et al.: Nanocomposite resist systems for next generation lithography. Microelectron. Eng. **63**(4), 391–403 (2002)
68. Cui, Z., et al.: Profile control of Su-8 photoresist using different radiation sources. Proc. SPIE. **4407**, 119 (2001)
69. Schneider, A., et al.: SU-8 resist and x-ray LIGA technique to produce 1mm high Ni gear wheel. In: Proc. 2nd International Workshop on Microfactories (2000)
70. Sarkar, M., Mohapatra, Y.N.: Electron beam lithography in thick negative tone chemically amplified resist: controlling sidewall profile in deep trenches and channels. Microelectron. Eng. **130**, 1–7 (2014)
71. Cui, Z., et al.: Comparative study of AZPN114 and SAL601 chemically amplified resists for e-beam nanolithography. J. Vac. Sci. Technol. **B16**(6), 3284 (1998)
72. Cui, Z., Prewett, P.D.: Comparative study of AZPF514 and UVIII chemically amplified resists for e-beam nanolithography. Microelectron. Eng. **46**(1–4), 255 (1999)
73. Nordquist, K.J., Resnick, D.J., Ainley, E.S.: Comparison of negative resists for 100 nm electron-beam direct write and mask making applications. J. Vac. Sci. Technol. **B16**(6), 3289 (1998)
74. Cui, Z., et al.: Optimised process for electron beam nanolithography using AZPN114 chemically amplified resist. Microelectron. Eng. **35**, 145 (1997)
75. Cui, Z., et al.: Sub-150nm electron beam lithography using chemically amplified resists. Proc. SPIE. **3049**, 667 (1997)
76. Yang, D., et al.: Performance of a high resolution chemically amplified electron beam resist at various beam energies. Microelectron. Eng. **155**, 97–101 (2016)
77. Maeda, N., et al.: Resist thickness dependence of line width roughness of chemically amplified resists used for electron beam lithography. Jpn. J. Appl. Phys. **59**, 086501 (2020)
78. Hu, S., et al.: Chemically amplified resist based on dendritic molecular glass for electron beam lithography. Chem. Res. Chinese Universities. **39**(1), 139–143 (2023)
79. Chen, W., Ahmed, H.: Fabrication of 5–7nm wide etched lines in silicon using 100keV electron-beam lithography and polymethylmethacrylate resist. Appl. Phys. Lett. **62**(13), 1499 (1993)
80. Broers, A.N.: Resolution limits for electron-beam lithography. IBM J. Res. Develop. **32**(4), 502 (1988)
81. Hasko, D.G., Yasin, S., Mumtaz, A.: Influence of developer and development conditions on the behavior of high molecular weight electron beam resists. J. Vac. Sci. Technol. **B18**(6), 3441 (2000)
82. Ocola, L.E., Stein, A.: Effect of cold development on improvement in electron-beam nanopatterning resolution and line roughness. J. Vac. Sci. Technol. **B24**(6), 3061 (2006)
83. Ocola, L.E., et al.: Lithography for sub-60 nm resist nanostructures. J. Vac. Sci. Technol. **B17**(6), 3164 (1999)
84. Yang, H., et al.: Comparative study of e-beam resist processes at different development temperature. Microelectron. Eng. **84**, 1109–1112 (2007)

85. Chen, Y., Yang, H., Cui, Z.: Effects of developing conditions on the contrast and sensitivity of hydrogen silsesquioxane. Microelectron. Eng. **83**, 1119–1123 (2006)
86. Tatarintsev, A.A., et al.: The effect of temperature on the development of a contrast HSQ electronic resist. Russ. Microelectron. **49**, 151–156 (2020)
87. Camino, F.E., Manfrinato, V.R.: Single-digit nanometer electron-beam lithography with an aberration-corrected scanning transmission electron microscope. J. Vis. Exp. **139**, e58272 (2018)
88. Manfrinato, V.R., et al.: Resolution limits of electron-beam lithography toward the atomic scale. Nano Lett. **13**, 1555–1558 (2013)
89. Vieu, C., et al.: Electron beam lithography: resolution limits and applications. Appl. Surf. Sci. **164**, 111–117 (2000)
90. Pease, R.F., et al.: Prospects for charged particle lithography as a manufacturing technology. Microelectron. Eng. **53**, 55–60 (2000)
91. McGrath, D.: The horror of 80-hour photomask write times. EE Times, 23 September 2010
92. Ronse, K.: E-beam mask-less lithography: prospects and challenges. Proc. SPIE. **7637**, 76370A-1 (2010)
93. Nakayama, Y., et al.: Electron-beam cell projection lithography: a new high-throughput electron-beam direct-writing technology using a specially tailored Si aperture. J. Vac. Sci. Technol. **B8**(6), 1836 (1990)
94. Pfeiffer, H.C., Stickel, W.: PREVAIL- an e-beam stepper with variable axis immersion lenses. Microelectron. Eng. **27**, 143 (1995)
95. Harriott, L.R.: SCALPEL: projection electron beam lithography. In: Proceedings of the 1999 Particle Accelerator Conference, IEEE, New York (1999)
96. WikiChip. https://en.wikichip.org/
97. Pfeiffer, H.C.: Recent advances in electron beam lithography for high-volume production of VLSI devices. IEEE Trans. Electron Devices. **ED-26**(4), 663 (1979)
98. Pfeiffer, H.C.: Variable spot shaping for electron-beam lithography. J. Vac. Sci. Technol. **15**, 887–890 (1978)
99. JEOL. https://www.jeol.com/
100. Vistec. https://www.vistec-semi.com/
101. NuFlare. https://www.nuflare.co.jp/
102. Pang, L.: Inverse lithography technology: 30 years from concept to practical, full-chip reality. J. Micro/Nanopattern. Mater. Metrol. **20**(3), 030901 (2021)
103. Christof Klein, E.P.: MBMW-101: world's 1st high-throughput multi-beam mask writer. Proc. SPIE. **9985**, 998505 (2016)
104. Platzgummer, E.: Maskless lithography and nanopatterning with electron and ion multi-beam projection. Proc. SPIE. **7637**, 763703–763701 (2010)
105. Petric, P., et al.: Reflective electron beam lithography: a maskless ebeam direct write lithography approach using the reflective electron beam lithography concept. J. Vac. Sci. Technol. B. **28**(6), C6C6 (2010)
106. Rio, D., et al.: 5 kV multielectron beam lithography: MAPPER tool and resist process characterization. J. Vac. Sci. Technol. B. **28**(6), C6C14 (2009)
107. Multi-Beam Mask Writer (MBMW). https://www.ims.co.at/
108. Platzgummer, E., et al.: Electron multibeam technology for mask and wafer writing at 0.1 nm address grid. J. Micro/Nanolithogr. MEMS MOEMS. **12**(3), 031108 (2013)
109. Klein, C., et al.: Projection mask-less lithography (PML2). Microelectron. Eng. **87**, 1154–1158 (2010)
110. Yasuda, J., et al.: Recent progress and future of electron multi-beam mask writer. Jpn. J. Appl. Phys. **62**, SG0803 (2023)
111. Willis, J.: Multi-beam mask writers are a game changer. In: Semiconductor Engineering (22 December 2022)

Chapter 4
Nanofabrication by Focused Ion Beam



4.1 Introduction

Ions are different from electrons because of their heavy mass. A hydrogen ion is 1840 times heavier than an electron. The heavy mass of ions makes them particularly suitable for nanofabrication by direct structuring of material, rather than by exposure of a polymer resist as e-beam lithography does. Ions were discovered and made use of in many applications other than nanofabrication, and many years earlier than electrons. The application of ion beam has nearly a hundred years of history. Since Thomson developed gas discharge ion source in 1910, ion beam technology was mainly used in material analysis, separation of isotopes, and modification of material property by ion implantation. In the early days, fabrication utilizing ion beam in nanometer scale was hampered because a fine size of ion beam was difficult to achieve from the early gas discharge ion sources. Focused ion beam (FIB) really became a nanofabrication tool following the introduction of liquid metal ion sources (LMIS) [1]. The LMIS made it possible to focus an ion beam to less than 5 nm in beam size. However, there is a drawback of using LMIS: the metallic ions in a LMIS can implant into a processed material. Plasma ion sources using xenon or argon gases were then developed. With properly designed ion optical system, sub-100 nm size of focused ion beam can be achieved. For even smaller beam size, there is the gas field ion source (GFIS) which can produce sub-1 nm beam, enabling even higher resolution than a LMIS for nanofabrication. On the one hand, a focused ion beam can machine a material by sputtering away the material. On the other hand, a focused ion beam combined with chemically active gases can deposit materials onto a substrate. A focused ion beam can also implant different atomic species into a substrate by using ion sources consisting of different elemental materials. All the above processes are based on the high resolution capability of a FIB system, which enables the processed structures to be in micrometer and nanometer dimensions. FIB system is truly a flexible and versatile nanofabrication tool.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_4

4.2 Ion Sources

An ion source is at the heart of a FIB system. Early ion sources were developed for mass spectrometry and nuclear physics research. Since the 1960s, ion implantation application for semiconductor manufacturing further pushed ion source development. It was the demand for nanoscale structuring that pushed development of various ion sources to be suitable for focused ion beams. Currently there are three types of ion sources being used for FIB applications: liquid metal ion source, plasma ion source, and gas field ion source.

4.2.1 Liquid Metal Ion Source

Liquid metal ion source (LMIS) is based on ion emission from liquid metal under a strong electrical field. LMIS was developed as a by-product when Clampitt and his colleagues were studying cesium ion sources for satellite thrusters in the 1970s [2]. The basic structure of a LMIS is illustrated in Fig. 4.1. A tungsten wire of about 0.5 mm diameter is electrochemically etched into a needle with the tip radius of only 5–10 μm. The tungsten needle is then wetted with molten metal. A small amount of liquid metal adheres to the needle. When applied to a high voltage, an electrostatic

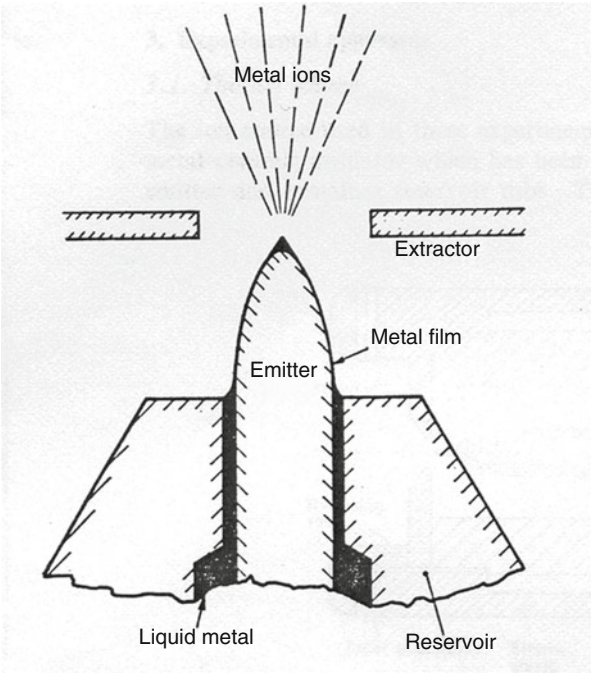


Fig. 4.1 Schematic of liquid metal ion source structure

force exerts to the liquid metal, pulling it into an extremely small apex (only about 5 nm in size). The electrical field strength at the liquid apex can be as high as 10^{10} V/m. At such extremely high field, metal atoms at the liquid apex become ionized and escape from the liquid metal surface in the form of field evaporation, resulting in ion emission. Although the total emission current may be only a few microamperes, the current density can be as high as ~10^6 A/cm^2 because of the extremely small emission area. The angular intensity of ion emission is up to ~20 μA/sr [3].

The ion emission process in a LMIS is a very complex dynamic process. For field electron emission discussed in Sect. 3.2.2 of Chap. 3, the electrons emit from a solid surface, while for ion emission in a LMIS, the emission surface is a liquid metal tip which changes shape with electrical field and emission current. In addition, the liquid metal tip has to have a continuous supply of material in order to compensate the loss of liquid metal due to emission. The entire emission process is the interplay between electrohydrodynamics of liquid metal and field ion emission. There had been analytical model [4] and numerical model [5] in attempt to describe such a dynamic process. The analysis indicated that there are three conditions which have to be satisfied in order to maintain a stable ion emission in a LMIS [6]:

1. The emission surface must have a certain shape in order to establish surface electrical field.
2. The surface electrical field must be high enough to maintain a certain level of ion emission current and a certain rate of liquid metal supply.
3. The liquid metal flow rate must be enough to compensate the loss of material corresponding to the ion emission current so that the stable shape of emission surface is maintained.

The three conditions have reflected the interdependence of LMIS emission surface shape, electrical field, emission current, and liquid metal flow. In a practical LMIS, the most critical condition for a sustained stable emission is to ensure a good wetting of tungsten needle by the liquid metal. A well-wetted needle can maintain a continuous liquid metal film over the needle surface, which not only helps to establish a liquid emission tip but also guarantees constant supply of liquid metal material to the tip.

Experiments have revealed the following characteristics of LMIS:

1. There exists a threshold voltage at which the onset of ion emission occurs. The threshold voltage depends on liquid metal property, radius, and sharpness of supporting tungsten needles. It is averagely above 2 kV. Once applied voltage exceeds the threshold value, the emission current increases rapidly.
2. The emission angle from a LMIS is relatively large, which is generally around 30° based on various experimental measurements. The higher the emission current, the larger the emission angle. However, the usable current will decrease with the increase of emission angle.
3. The angular distribution of emission current is relatively uniform.

4. The energy spread of emitted ions is relatively large, depending on the emission current. The higher the emission current, the larger the energy spread [7]. Large energy spread will cause large chromatic aberration in ion optical system.
5. Mass spectrometry analysis of LMIS emission has shown that at low emission current (≤10 μA), the emission consists of almost 100% of singly charged ions. With the increase of emission current, the proportion of multiple charged ions, charged molecules, ion clusters, and charged liquid droplets in total emission current increases, which has detrimental effects for the application of FIB technique.

The above characteristics indicate that for practical FIB applications, it is not desirable for a LMIS to work at large emission current condition. The optimum working current should be less than 2 μA. At small emission current, both the ion energy spread and emission angle are small, resulting in efficient use of emitted ions. The actual usable ion current is not necessarily lower than working at large emission current.

LMIS started with gallium as the emission material because the melting point of gallium is at 29.8 °C. Liquid gallium can adhere to tungsten surface very well. Gallium is presently still the source material for majority of LMIS used in FIB systems. Apart from gallium, LMIS made of other materials had also been investigated. The materials that had been tried for LMIS were Al, As, Au, B, Be, Bi, Cs, Cu, Ge, Fe, In, Li, Pb, P, Pd, Si, Sn, U, and Zn [8]. Some of the above elements can be directly made into single material LMIS. Some have to be made into eutectic alloys which are of low melting point. For liquid metal alloy ion source (LMAIS), the different ion species can be separated by magnetic field. Among those elements, As, B, Be, and Si are of particular interests because they are the main doping elements for semiconductor transistors. In the early years when LMIS-based FIB technology was developed, alloy LMIS consisting of these elements were used for direct ion implantation to make transistors. Though LMAIS was no longer used to make transistors, there are renewed interests in LMAIS recently, mainly because ion species other than gallium can avoid the gallium staining problem which will be introduced later in this chapter, as well as adding special functionalities to a FIB fabricated structure [9].

The lifetime of a LMIS is measured by μA•hour. It is obvious that the larger the emission current, the shorter the lifetime of a LMIS. Current commercial gallium LMIS has lifetime up to 6000 μA•hour. The limiting factors for longer lifetime include not only the supply of liquid metal material but also the contamination caused by backscattered substrate materials which redeposit onto the LMIS emission surface. Such redeposition of alien materials can change the flow characteristics of liquid metal. In particular, when a FIB is used for chemical gas-assisted etching or deposition, some organic molecules can absorb onto liquid metal surface. Upon bombardment of secondary electrons, these organic molecules can form a carbon layer, causing significant increase in flow resistance of liquid metal [10]. Some LMIS using high vapor pressure metals are also of short lifetime because of their volatility at molten state.

4.2.2 *Plasma Ion Source*

Ion species generated by a plasma have long been used in mass spectrometry. They were not used for FIB applications because of low current density and large energy spread, which is difficult to focus into a fine beam. The renewed interest in plasma ion source for FIB started with the emergence of inductively coupled plasma (ICP). Figure 4.2 shows the schematic of ICP ion source. With the assistance of induction coil outside the plasma chamber, the plasma density inside the chamber is greatly enhanced. The density of Xe^+ plasma can be as high as 10^{13} cm^{-3} [11]. The ions inside the chamber are extracted by the extraction electrode at 4 kV and accelerated by an ion optical column to 20–30 keV energy. Plasma ion source is a broad source and the source size is the extraction aperture (~200 μm) which cannot be too small. Therefore, even with a focusing lens, the final spot size at the substrate target is much larger than the ion beam from a LMIS, ~50 nm compared to ~5 nm of LMIS [12]. Also due to the source size, the current density is two orders of magnitude lower than LMIS, ~10^4 A/cm^2 compared to ~10^6 A/cm^2 of LMIS. However, the angular intensity of ICP source is much higher than a LMIS, ~10 mA/sr compared to ~15 μA/sr of LMIS, which makes the ICP source much advantageous in high current operation [11]. Using a 1 μA Xe^+ beam can have 80 times higher sputtering removal rate for silicon than a Ga beam operating at 20 nA [13]. Another advantage of ICP source compared to LMIS is it being free from Ga staining, because it uses gas species such as Xe, Ar, or O_2.

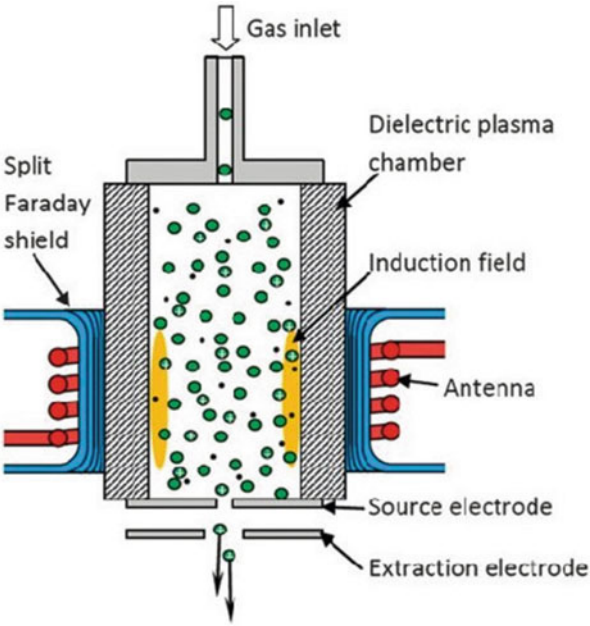


Fig. 4.2 Schematic of inductively coupled plasma (ICP) ion source. (Reprinted from [11] with permission)

4.2.3 *Gas Field Ion Source*

While the plasma ion source avoids using heavy metal ion species but does not have a fine beam size, a gas field ion source (GFIS) emits gaseous ions but has extremely small beam size. Gaseous atoms, mostly helium atoms, absorbed on the surface of a sharp needle can be ionized by the extremely high electrical field at the apex of needle, very much like the LMIS where Ga atoms are field-ionized. Field-induced ion emission was discovered more than 70 years ago and has been utilized ever since as field ion microscope (FIM) to observe atomic structure of needle materials [14]. The advance of FIM to become GFIS for FIB applications originated from the work reported by Ward et al. in 2006 [15]. A conventional FIM will have many gas atoms absorbed on the emitter tip. Ward et al. managed to control the electrical field and gas inlet so that only three atoms are absorbed and field-ionized on the tip apex, as illustrated in Fig. 4.3 [16]. The extraction field is then configured to allow only one ionized atom to produce the beam while the other two are dumped [15]. Such an ion source can produce the ultimate focused ion beam size of 0.25 nm, energy spread <0.5 eV, and angular current density of 2.5 μA/sr. With such a small beam spot, the source brightness is estimated to be ~10^9 A/cm^{2}sr, which is three orders of magnitude better than a gallium LMIS, and many nanofabrication challenges can be met by helium FIB [17]. Apart from helium, other heavier inert gases such as neon [18] and xenon [19] have also been used in GFIS.

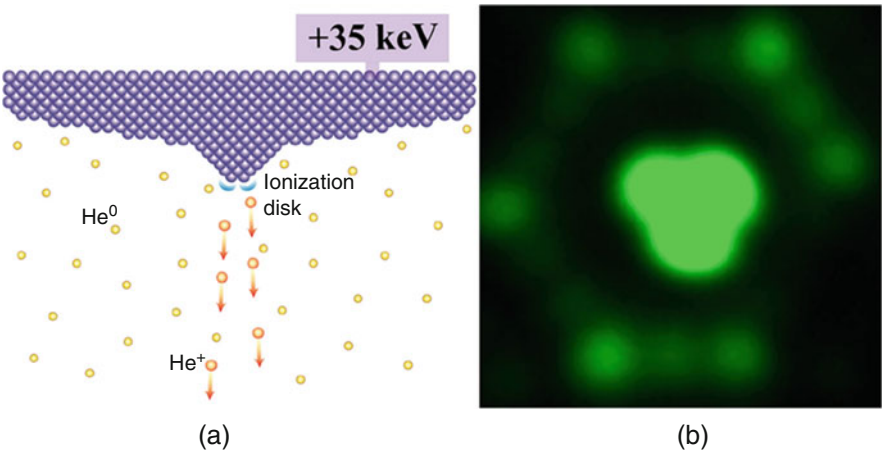


Fig. 4.3 (**a**) Schematic of helium atom absorption and ionization, (**b**) field ion image of three ionized helium atoms. (Reprint from [16] with permission)

4.3 Focused Ion Beam Systems

Focused ion beam (FIB) systems have no fundamental difference from electron beam lithography systems. They all consist of emission source (electron or ion source), lens column (electron or ion optics), work stage, vacuum, and control systems. The core part of focusing emitted ions into a fine beam is the ion optical system. Electron optics has been introduced in Chap. 3. Ion optics is very much like electron optics. The only difference is that ions can have different masses and charges. The force exerted on charged particles in electrical field depends only on the charge a particle carries, as shown in Eq. (4.1a), whereas the force exerted on charged particles in magnetic field depends not only on charge but also on mass, as indicated by Eq. (4.1b).

$$F_e = qE \tag{4.1a}$$

$$\vec{F}_m = q\vec{v} \times \vec{B}, \quad v = \sqrt{\frac{2qV}{m}} \tag{4.1b}$$

where ν is velocity and V is acceleration voltage.

Electrical charge over mass is commonly called charge-mass ratio. It is apparent that ions with different charge-mass ratio will experience magnetic force differently in a magnetic field. Such a property is useful for some applications but detrimental for some other applications. For example, mass spectrometry utilizes this feature to select ions of different mass in a magnetic field. For FIB application, ions of different mass will be focused to different focal planes. Therefore, FIB systems are universally using electrostatic lens and deflectors, which have been described in Chap. 3, to focus and deflect ion beam.

A typical FIB system is schematically shown in Fig. 4.4a. The focusing lenses are electrostatic lenses. Ion beam deflector is an electrostatic octopole deflector. The typical acceleration voltage of FIB system is between 1 and 30 kV, which is also the energy the ion beam carries to sputter at a target substrate. The working current is different from the emission current of an ion source. The typical working current is 0.2 pA to 50 nA for a Ga^+ LMIS, 0.1 pA to 10 μA for a Xe^+ ICP source, and 0.1 pA to 10 pA for a He^+ GFIS [11], which is adjusted by using an aperture.

There are a number of commercial FIB system manufacturers worldwide. FEI Company was the first to use LMIS to build commercial FIB systems. The company was acquired by Thermo Fisher Scientific Inc. in 2017. Most of the commercial FIB systems are not stand-alone FIB but combined with SEM (scanning electron microscope). A schematic of dual-beam system is shown in Fig. 4.4b, which enables in situ observation of FIB sputtering or deposition progress. Thermo Fisher Scientific offers Helios series dual-beam systems which use either a gallium LMIS or a plasma ion source [21]. Other commercial dual-beam systems include the VELION from Raith [22], NX9000 from Hitachi [23], and JIB systems from JEOL [24]. Currently,

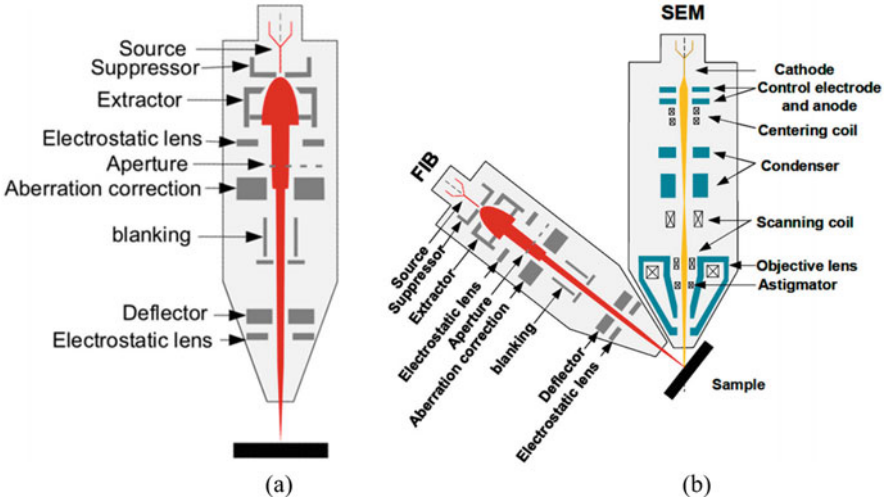


Fig. 4.4 Schematics of (**a**) FIB system, (**b**) FIB-SEM dual-beam system [20]

only ZEISS offers the GFIS-based FIB system: the ORION NanoFab system [25], which not only has a GFIS with helium and neon ions but also has a gallium LMIS to complement.

4.4 Ion Scattering in Solid Materials

As charged particles, high energy ions also scatter in solid material. The difference between electron scattering and ion scattering is that ions are much heavier than electrons; therefore, they scatter very little. Ion scattering can also be classified as elastic scattering and inelastic scattering. Elastic scattering changes the direction of incident ions without loss of energy. Inelastic scattering causes loss of energy of incident ions. The scattering angle can be calculated by nuclear differential scattering cross section expressed by Eq. (4.2) [26]:

$$\mathrm{d}\sigma = \frac{\pi a^2}{2} t^{\frac{3}{2}} f\left(t^{\frac{1}{2}}\right) \mathrm{d}t \tag{4.2}$$

where $t^{\frac{1}{2}} = \varepsilon \sin\left(\frac{\theta}{2}\right)$; $\varepsilon = \frac{aM_2}{e^2 Z_1 Z_2 (M_1 + M_2)} E_0$; $a = \frac{0.8853 a_0}{\left(Z_1^{\frac{2}{3}} + Z_2^{\frac{2}{3}}\right)^{\frac{3}{2}}}$; $a_0 = 0.529A°$

$f\left(t^{\frac{1}{2}}\right) = \lambda t^{\frac{1}{2}-m}\left[1 + (2\lambda t^{1-m})^q\right]^{-\frac{1}{q}}$; $\lambda = 1.307$, $m = 0.333$, $q = 0.667$

Z_1 and M_1 are atomic number and atomic weight of incident ions. Z_2 and M_2 are the atomic number and atomic weight of solid material atoms. E_0 is the initial energy of incident ions. As mentioned in Chap. 3, collision of electrons with atoms is elastic

scattering in which electrons do not lose energy because electrons are far smaller than atoms. Ions colliding with atoms in a solid material will lose part of their energy because of their comparable masses.

There are two types of energy losses during ion scattering in solid. One is nuclear losses, i.e., an incident ion scattering with atomic nucleus transfers part of its energy to the solid atom, causing the atom to either dislocate its position to become recoiling atoms or escape from the solid surface. The phenomenon of incident ions knocking solid atoms out of solid target is called sputtering. Another is electronic losses, i.e., the incident ion transfers part of its energy to the electrons surrounding an atomic nucleus. These electrons can either be excited to produce secondary electron emission from solid surface or be stripped off from the atom, resulting in the ionization of atom and secondary ion emission. Atomic nucleus losses can be calculated by Eq. (4.3), and electronic losses can be calculated by Eq. (4.4) [27], where L is the mean free path between two collisions and N is the atomic density of solid material.

$$\Delta E_n = \frac{4M_1M_2}{(M_1+M_2)}E_0 \sin\left(\frac{\theta}{2}\right) \tag{4.3}$$

$$\Delta E_e = LNS_e(E_0) \tag{4.4}$$

with $S_e = K_L\sqrt{E_0}$ and $K_L = 1.211\dfrac{Z_1^{\frac{7}{6}}Z_2}{\left(Z_1^{\frac{2}{3}}+Z_2^{\frac{2}{3}}\right)^{\frac{3}{2}}\sqrt{M_1}}$.

Like Monte Carlo simulation of electron scattering introduced in Chap. 3, ion scattering can also be simulated by Monte Carlo method. The simulation process is similar, i.e., the total scattering cross section is calculated by integration of Eq. (4.2); the probability of scattering with a particular atomic species is determined by the ratio of their total scattering cross sections; scattering angle and mean free path of scattering are chosen randomly; atomic nucleus and electronic losses are calculated by Eqs. (4.3) and (4.4); the ion scattering trajectory is followed until it loses all the energy and stops at some location inside the solid material. Figure 4.5 shows the

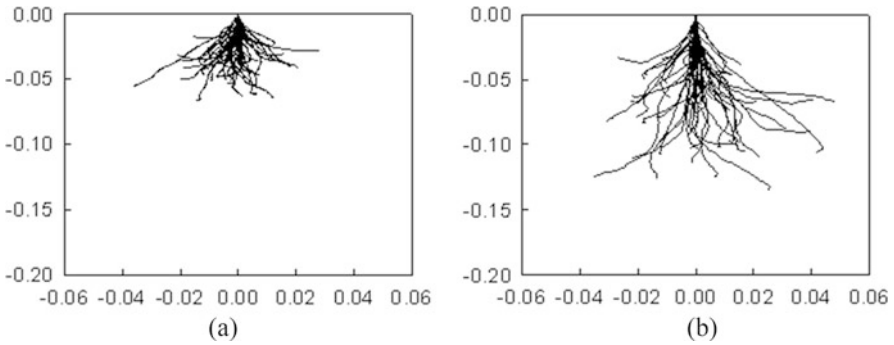


Fig. 4.5 Gallium ion scattering trajectories calculated by Monte Carlo method, (**a**) ion energy = 50 keV, (**b**) ion energy = 100 keV (dimension unit: μm)

gallium ion scattering trajectories in silicon, calculated by Monte Carlo method [28]. As it is shown in Fig. 4.5a, 50 keV gallium ions can only penetrate in silicon up to 50 nm deep because of its heavy mass. Even at 100 keV energy, the penetration is only around 100 nm as shown in Fig. 4.5b. Compared with electron scattering, electrons with the same energy and in the same substrate material have scattering depth of 10 μm. The ion penetration depth is also called stopping range. The stopping range increases with the increase of ion energy and the decrease of ion mass. Unlike electron scattering, incident ions will eventually rest in the solid material after losing all their energy and become part of the material. This is the so-called ion implantation. Ion implantation can modify the material property and will be described in later section.

For Monte Carlo analysis of ion scattering and sputtering, there is a free downloadable computer program "SRIM" (Stopping and Range of Ions in Matters) [29]. The program was initially called "TRIM" [30]. Since published in 1980, it has been upgraded several times [31]. The latest version is SRIM-2013. The program can calculate 3D distribution implanted ions in solid, damage of target material caused by ion bombardment, sputtering, generation of secondary ions, phonon production by ion excitation, etc. The target material can be single material or composite material consisting of up to eight different materials. Sputtering yield and ion scattering trajectories can be displayed during simulation.

The above analysis has assumed that the target material is amorphous material where atoms are randomly arranged. The scattering of incident ions in an amorphous material is a random process. If the target material is a crystalline material where atoms are orderly arranged, incident ions may penetrate much deeper along the low indexed axis of crystalline. The penetration depth can be several times of that for amorphous target. Such a phenomenon is called ion channeling effect. In this case, ion energy loss is mainly electronic loss, and the yield of sputtered target atoms is low because of low probability of incident ions colliding with target atoms. This can be a serious issue when sputtering a polycrystalline material with ion beam. Because different crystalline structures coexist in a polycrystalline material, there is different indexed axis on the target surface, leading to different sputtering yields at different locations. The final sputtered surface can be very rough, as shown in Fig. 4.6a which is the surface topography of sputtered copper layer by FIB. Copper on semiconductor substrate is always in the form of polycrystalline structure. The highly rough surface after FIB sputtering is caused by differentiated sputtering yield across the surface. One way to reduce the sputtering roughness is to deliberately block the ion channels by doping the copper layer with impurity atoms. The doped copper layer will be like amorphous layer and sputtering uniformity can be greatly improved, as shown in Fig. 4.6b [32].

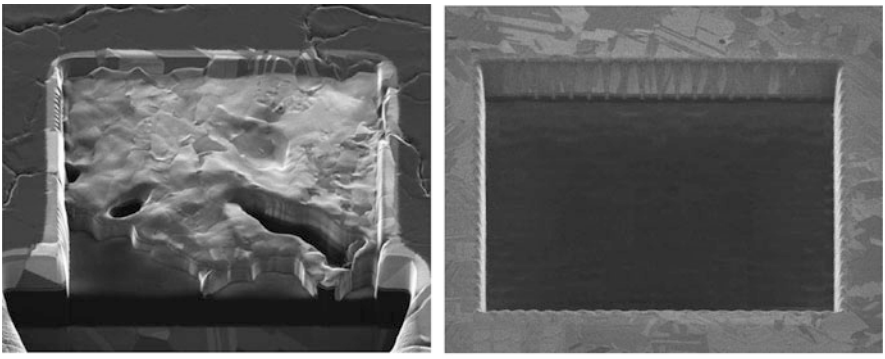


Fig. 4.6 FIB sputtering of polycrystalline copper, (**a**) with channeling effect (**b**) after doping

4.5 Ion Sputtering and Deposition

Because of the heavy mass of ions, a FIB tool can do much more than e-beam lithography in nanoscale fabrication. The main functions of FIB are sputtering (removing materials) and deposition (adding materials). These are direct nanofabrication, without having to expose resist first and etching or deposition subsequently, like the electron beam lithography introduced in Chap. 3.

4.5.1 Ion Sputtering

Ion sputtering is a process where incident ions transfer their energy to atoms in a solid material, and upon acquiring enough kinetic energy from incident ions, these atoms escape from solid material surface. Target materials can be removed layer by layer with ion sputtering, which is also called ion milling. Ion sputtering is not a one to one process. An ion beam bombarding a target material can generate a large number of recoiling atoms. These recoiling atoms can further pass their kinetic energy to surrounding atoms, generating more recoiling atoms. Some of the recoiling atoms at the vicinity of target surface may have enough kinetic energy to break off from the binding of surface energy into free space, becoming sputtered atoms. The most important parameter of ion sputtering is sputtering yield, i.e., the number of sputtered atoms every incident ion can generate. The sputtering yield can be calculated by Sigmund's "linear collision cascade" model, as shown in Eq. (4.5) [33].

Table 4.1 Ion sputtering yields of various target materials by Ga, Al and In ions

	Ga^+		Al^+		In^+	
Target	*Y*	*R*	*Y*	*R*	*Y*	*R*
Si	2.1	0.26	0.9	0.11	2.9	0.36
Cu	1.8	0.13	0.9	0.07	5.2	0.39
Zr	2.5	0.36	1.2	0.18	1.9	0.28
Mo	1.1	0.10	0.8	0.07	2.0	0.19
Ta	3.5	0.40	1.3	0.15	2.0	0.22
W	1.1	0.11	0.8	0.08	2.1	0.21

Y sputtering expressed by number of atoms per incident ion
R sputtering yield expressed by volume (μm_3) per nC (10^{-9} C) ion dose

$$Y(E,\theta) = \frac{0.042}{U_s}\alpha S_n(\varepsilon)\cos^{-f}(\theta) \tag{4.5}$$

The model predicts that ion sputtering yield depends not only on incident ion energy but also on incident angle, atomic density, and mass of target material. Although ion sputtering yield can be calculated theoretically, most of sputtering yield data are experimentally measured at specific sputtering conditions. Table 4.1 gives the sputtering yields for some of the target materials with gallium, aluminum, and indium ions at 30 keV initial energy [34]. A collection of sputtering yields for different target materials using gallium FIB of different energies can be found in [35].

Sputtering yield *Y* is a constant and only related to ion species and energy, target material property, and sputtering angle. Once the sputtering yield is known, the actual sputtering rate can be calculated by multiplying *R* in Table 4.1 with ion dose. For example, the sputtering yield for aluminum target is known at 0.3 μm^3/nC, and the aluminum sputtering rate will be 0.03 μm^3 per second for an ion beam current of 100 pA. If a 5 × 5 μm hole with 5 μm depth is to be machined by FIB on an aluminum substrate, the total volume of removed material is 125 μm^3. According to the above sputtering rate, the rate of sputtering depth is 7.2 nm per minute, and it needs 69.4 min to machine the 5 × 5 × 5 μm hole. The total sputtering time can be changed by changing the ion beam current.

From theoretical analysis and experimental investigation, it is revealed that the ion sputtering yield has the following characteristics:

1. Ion sputtering yield rises with glancing angle of beam incidence and reaches its maximum at about 80°. Further increase of the angle toward 90° results in rapid drop of sputtering yield. The sputtering yield versus incident angle is plotted in Fig. 4.7, which indicates that sputtering at vertical incidence of beam (0°) does not generate the maximum sputtering yield. This is because vertically incident ions can penetrate much deeper due to the channeling effect, which produces less recoiling atoms, hence low sputtering yield.
2. High energy ions do not necessarily generate high sputtering yield. For gallium ions, the sputtering yield becomes saturated above 30 keV ion energy, which is



Fig. 4.7 Typical relation between sputtering yield and ion beam incident angle

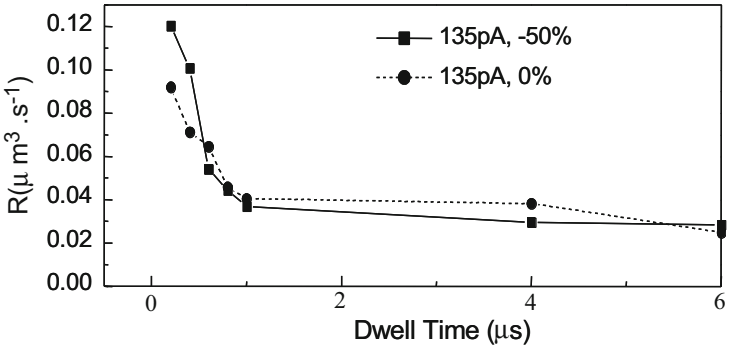


Fig. 4.8 Variation of sputtering rate with ion beam dwelling time

one of the reasons that most commercial FIB systems work at 10–30 keV beam energy.

3. Redeposition of sputtering material can reduce the net sputtering rate. At a fixed beam current, a slow scan appears to have high sputtering rate because more ions are injected into target; in reality slow scan reduces the sputtering rate as the experimental results of the sputtering rate vs. ion beam dwelling time shown in Fig. 4.8 [36]. Longer dwelling time (slow scan) does not produce higher sputtering yield because of the redeposition effect. There is a saturation point beyond which the sputtering removal of material and redeposition reach a balance. It is found in experiment that the sputtering yield is a constant for ion beam scanning speed varying from 0.05 cm/s to 1.0 cm/s [37]. This phenomenon

Table 4.2 Sputtering yield enhancement factors

Gases	Al	Si	SiO_2	W
Br_2	8–16	5–6	0	0
Cl_2	7–10	0	0	0
XeF_2	0	7–12	7–10	7–10

is most significant when sputtering deep holes with FIB. As the hole depth increases, more and more sputtering atoms redeposit on the sidewalls of the hole. One method to reduce the redeposition is to reduce the dwelling time of an ion beam at each sputtering point, for example, to sputter with fast and multiple scans of ion beam.

4. Ion bombardment combined with chemically active gases can greatly enhance the sputtering yield. The chemical gas-assisted sputtering process is essentially a combination of ion sputtering and chemical gas phase etching, because ionized chemically active gases such as chlorine and fluorine can violently react with silicon, gallium arsenide, or aluminum to produce volatile compounds. The process works as the following: a small amount of chemically active gas, called precursor gas, is introduced into the sputtering chamber; gas molecules then absorb to target surface; ion bombardments ionize the absorbed gas atoms; these gas ions react with target surface material and convert the target material into volatile compounds which are then evacuated by the vacuum system of FIB. Experiments have revealed that chemical gas-assisted ion sputtering has sputtering rate of a few ten times higher than normal ion sputtering [38]. The enhancement factor can be expressed as Eq. 4.6 [36].

$$E = 1 + (E_0 - 1)\theta \tag{4.6}$$

where $\theta = \dfrac{1}{1 + \frac{\phi J}{F}}$

In Eq. 4.6, θ is the gas molecule coverage over the surface ($\theta < 1$); J is the ion current density; ϕ is related to absorption property of material surface; F is the flux of reaction gas flow; and E_0 is enhancement factor at maximum coverage of gas molecules ($\theta = 1$). Table 4.2 lists the enhancement factors for some combinations of reaction gases and sputtered materials. Eq. 4.6 indicates that the enhancement factor is related to the gas molecule coverage which is directly proportional to the gas flux and inversely proportional to ion current density. It must be pointed out that gas-assisted sputtering only works for some specific combinations of materials and gases. These materials have to be reactive to the chemical gases.

The most effective technique to reduce the aforementioned redeposition is to incorporate chemical gases into the sputtering process. Figure 4.9 compares 50 nm interdigital finger structures patterned by ion sputtering etching without gas assistance and with gas assistance. Because of the tightly confined spaces between the fingers, they are barely resolvable by ion sputtering alone due to redeposition

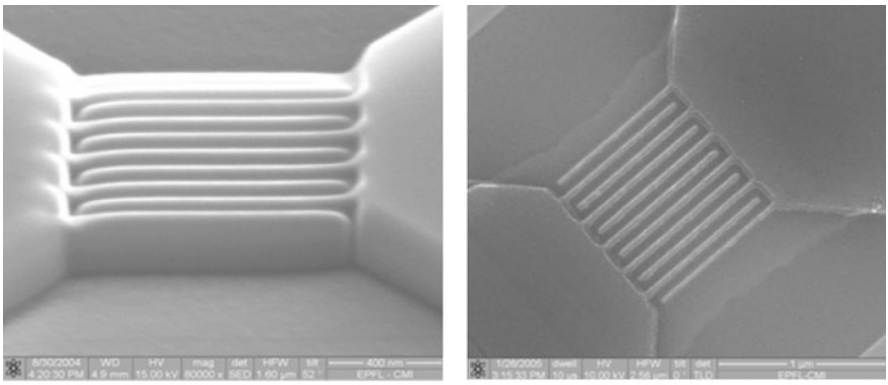


Fig. 4.9 Ion sputtering patterning of 50 nm interdigital fingers, (**a**) without XeF_2 gas assistance and (**b**) with XeF_2 gas assistance. (Reprint from [39] with permission)

(Fig. 4.9a). With the assistance of XeF_2 gas, the 50 nm fingers are clearly separated (Fig. 4.9b).

The above discussions are based on the assumption of using a FIB with gallium LMIS. There are also FIB systems based on ICP plasma ion source and gas field ion source (GFIS) as introduced in Sect. 4.3. The milling rate is higher when a plasma ion source FIB is used because of its high ion current; the milling accuracy is better when a GFIS-based FIB is used because of its high resolution capability. It has been shown that the milling rates of a Xe ICP-FIB can be at least 50 times higher than a Ga LMIS-FIB with same energy of ions (30 keV) due to the availability of higher current [40]. For He GFIS, the milling rate is naturally low because of its extremely low ion current. However, the milling rate can be improved by using heavier ions such as neon or argon. For example, a neon FIB can have the sputter yield of $\sim$1 atom per incident ion for Si and 4 atoms per incident ion for Cu. This is about $\sim$100$\times$ higher than that of helium FIB for the same energy. Though it is about 2$\times$ lower than gallium FIB, the neon beam size from the GFIS is $<$1.5 nm which is of higher resolution than gallium FIB [18].

When sputtering multilayer materials, end point detection is needed to stop ion sputtering once a layer of material has been sputtered off [37]. The end point detection is usually achieved by observing the contrast of secondary electron or secondary ion image of the sputtered surface. As different materials have different image brightness due to the difference in secondary electron or secondary ion yield, a comparison of image brightness can tell if a layer material has been completely removed. Another method is to employ a secondary ion mass spectrometer. By analyzing the variation of species of secondary ions in free space, one can tell if certain material has been completely removed and new material underneath the layer has started to appear in the mass spectrum.

4.5.2 *Ion Beam Assisted Deposition*

If the precursor gas introduced into sputtering chamber is not chemically reactive to the substrate material, the ions produced by ion beam bombardment on the absorbed gas atoms will form nonvolatile compounds, which cannot be removed away by evacuation but stay on the target surface. This is the ion beam assisted deposition. While material is deposited by ion sputtering of absorbed gas atoms, ion sputtering itself can also remove the as-deposited atoms. The deposition process and sputtering process coexist and compete with each other. In this case, only fine tuning of ion energy, ion dose, gas flow rate, and chamber pressure can adjust the deposition rate to be greater than sputtering rate, so that the net result is material deposition, not sputtering.

The purpose of ion beam assisted deposition is to make functional structures, such as metal interconnections in IC chips. To deposit metal ions, metal organic compound gases should be used. Some commonly used metal organic compound gases with different metal constituents and composition of deposited metal films are shown in Table 4.3. As can be seen, the deposited film is not pure metal, which implies that the conductivity of FIB deposited compounds will not be as good as pure metals. These impurities are mostly carbon and oxygen atoms because of the organic compounds used. Heating the sample can remove some of the impurities and improve the conductivity. FIB assisted deposition can also deposit nonmetal material such as SiO_2. SiO_2 can be deposited by introducing water vapor into the FIB sputtering chamber.

4.6 Focused Ion Beam Direct Nanofabrication

Focused ion beam is the most versatile tool for nanofabrication. Unlike the optical lithography or electron beam lithography which has to pattern a polymer resist first as the imaging layer, FIB can directly remove materials by sputtering, or add materials by gas-assisted deposition as well as ion implantation, all of which are in

Table 4.3 Commonly used metal organic compound gases and composition of deposited films

Organic metal compound gases	Sputtering ion	Ion energy keV	Deposits	Composition	Resistivity of deposit μΩ-cm	Reference
$Wu(CO)_6$	Ga^+	25	W:C:Ga:O	75:10:10:5	150	[41]
$C_7H_7F_6O_2Au$	Ga^+	40	Au:C:Ga	80:10:10	3	[42]
$(CH_3)_3Al$	Ga^+	20	Al:O:Ga:C	37:27:26:10		[43]
$C7H_{17}Pt$	Ga^+	35	Pt:C:Ga:O	45:24:28:3	70	[44]



Fig. 4.10 Nanofabrication performed by ion beam assisted deposition (top) and ion beam sputtering (bottom)

the form of patterned structure by a finely focused ion beam. Figure 4.10 shows such patterned structures made by both ion beam sputtering and ion beam assisted deposition. With this unique capability, many applications have been explored. It has become an indispensable tool in scientific research laboratories as well as in some industrial sectors. However, one should also be aware that high energy ions can implant into the target material. Gallium ions can cause staining problem [45], and gaseous ions can create bubbles at the subsurface of target [46].

4.6.1 Inspecting and Editing Integrated Circuits

High density integrated circuit (IC) chips contain up to billions of transistors with multilayer interconnections. It is hard to avoid mistakes in designing such complex circuit systems. Once a circuit design becomes an IC chip, it can no longer be altered even though mistakes such as circuit routing in design have been made. However, by using the sputtering and deposition functions of a FIB, one can connect or disconnect some parts of circuits on an IC chip. By altering the interconnection, errors in circuit design can be diagnosed, and correction of these errors (debugging) can be directly carried out on the chip by FIB. Figure 4.11a shows part of an IC chip that has been modified. There are sputtered cuts to disconnect some circuits and depositions of tungsten (W) wires to reconnect or bypass some circuits. There are even deposited insulator patches to allow crossover of FIB deposited conductive lines. Modern FIB systems can compare the entire IC design layout with actual IC chip image (SEM imaging in a dual-beam system) point to point and locate the error points as well as make corrections precisely to perform high precision circuit editing [47]. In addition to diagnosing and repairing design mistakes in IC chips, FIB technology can also help to diagnose problems that occurred in manufacturing processes, for example, cutting through the IC chips to look at the cross section of different processed layers,

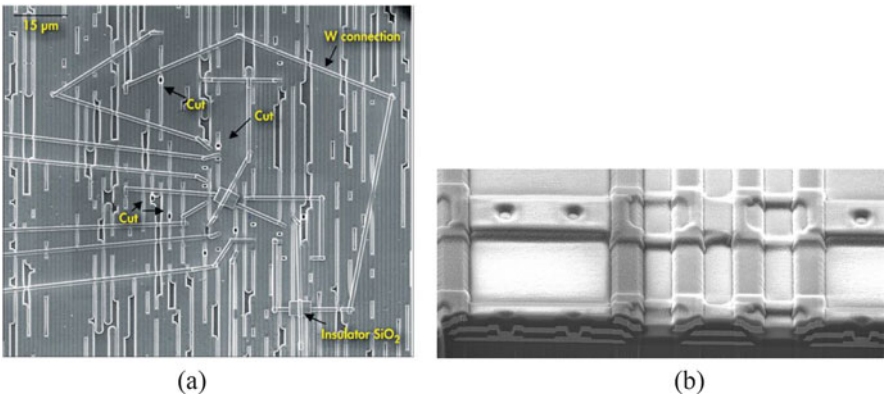


Fig. 4.11 (**a**) Debugging IC chip by FIB; (**b**) cross-sectioning IC chip by FIB

as shown in Fig. 4.11b which is a cross section of IC chip cut by FIB. If something went wrong in the process, one can find out where or in which process steps it went wrong by inspecting the cross section. The above examples demonstrated that the FIB is truly an indispensable tool in modern IC manufacturing. Nowadays, every wafer fab around the world is equipped with FIB tools to assist IC manufacturing.

Before the 22 nm node, the transistors in an IC chip are planar structures located at the bottom of circuit stack. FIB editing only involves modifying the multilayer interconnects on top levels. From 22 nm node onward, FinFET transistors became dominant, and at 5 nm node the gate-all-around (GAA) transistors have been introduced in ICs. The transistors are no longer planar and the thin dielectric layer capsulating the semiconductors is more susceptible to transistor performance. IC editing using a FIB needs to go deeper into the device and to carefully navigate around the more delicate dielectric layers. For the state-of-the-art circuit editing, the FIB tool works at lower landing energy of ions and lower beam current so as to reduce damage to sensitive circuits and allows the operator to remove very thin layers with surgical precision [48].

4.6.2 Repairing Defects of Optical Masks

Another major application of FIB is to repair defects in optical masks. Mask defects are the primary cause of circuit failures in an IC chip. These defects are generated during fabrication of optical masks. Optical masks are glass or quartz plates with chrome patterns made by lithography and etching. During lithography and etching processes, extra spots of chrome or missing spots of chrome may form to become opaque defects or clear defects, as illustrated in Fig. 4.12. These defects can be

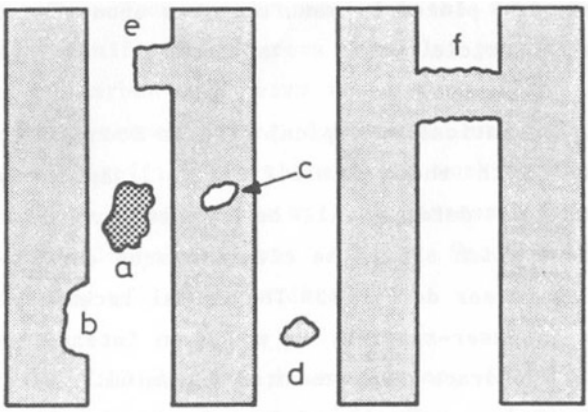


Fig. 4.12 Illustration of optical mask defects (opaque defects, **a**, **e**, **d**, **f**; clear defects, **b**, **c**)

imaged onto silicon wafer to become unwanted features in a circuit, leading to failure of the circuit.

For opaque defects, the repairing technique in early days was mainly by laser zapping. A high energy density laser beam heats the chrome defect, which vaporizes the extra chrome spot. However, laser beam spot is too big to repair small defects and the repaired region always has rugged edges. As IC feature dimension shrinks, FIB repair of mask defects became dominant. The principle of FIB repair of opaque defects is ion sputtering to remove unwanted chrome defects. The thickness of chrome layer on an optical mask is normally around 80 nm. FIB sputtering of chrome layer must be carefully controlled to avoid over sputtering. Extra sputtering after the chrome defect has been removed can result in new defect in the mask. End point detection is necessary in FIB repair of mask defects. A common method to detect the end point is by secondary electron image of repaired region, because of the different contrast in the secondary electron images between chrome and glass.

Repair of clear defects is relatively simple. It makes use of FIB assisted deposition process to patch up the clear defects usually with carbon material. Hydrocarbon gas is introduced into the vicinity of ion beam sputtering region by a small nozzle. Absorbent hydrocarbon molecules on target surface are dissociated upon ion beam bombardment, leaving carbon atoms at the target surface. Because high energy ions can also break up the bond of SiO_2 which is the mask substrate, the silicon atoms with dangling bonds at surface can form a strong bond with carbon atoms, holding the carbon atoms firmly to the surface. FIB assisted deposition of carbon can make a very robust repair of clear defects. The deposited carbon patches will not come off the mask substrate in normal mask cleaning process. As the same with FIB assisted deposition of other materials, gas flow rate, ion beam current, and dose must be precisely controlled and matched, so that gas molecules have enough time to be absorbed at the substrate surface. Ion beam sputtering time should be long enough to dissociate the absorbent molecules but not too long to over sputter the surface, causing removal of already deposited carbon atoms. With such a control, the deposition rate is higher than sputtering rate and continuous deposition of carbon

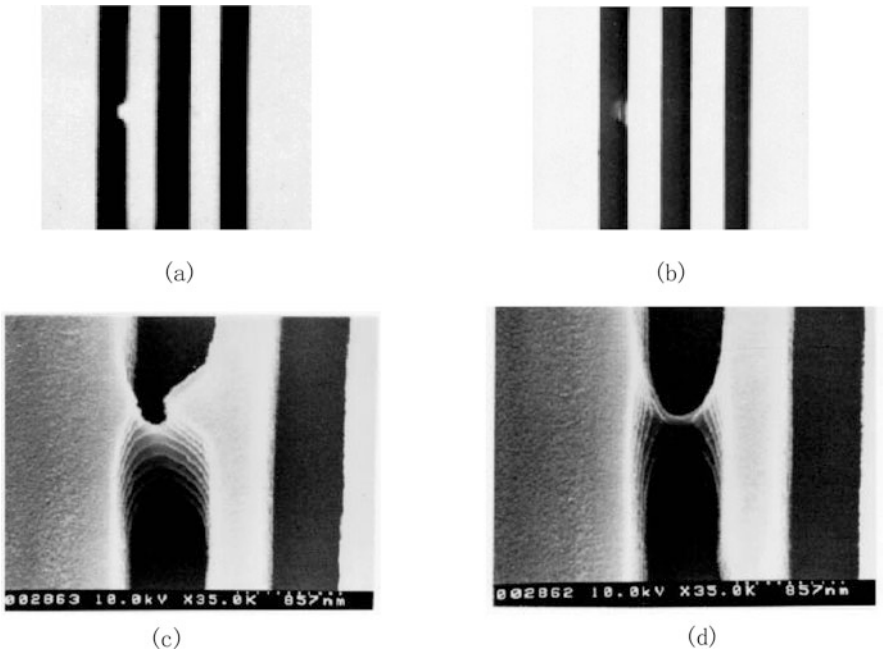


Fig. 4.13 (**a**) Photomask image with a defect, (**b**) photomask image after FIB repair, (**c**) photoresist image from the defected mask, (**d**) photoresist image from the repaired mask

is ensured. Clear defects are normally on glass or quartz surface which is an insulating material. Charge neutralization measure is needed to prevent ion charge accumulation.

The biggest problem with FIB repair of mask defects is gallium staining [45]. As mentioned earlier, gallium ions eventually remain in a target material after losing all energy to become implanted ions. As gallium is a heavy metal, implanted gallium ions can alter the light transmission of glass, which is called gallium staining. Although gallium ion sputtering can remove chrome opaque defects, the repaired region does not resume to its original transparency because of the gallium staining, resulting in incompletely repaired defects as compared in Fig. 4.13 before (Fig. 4.13a, c) and after (Fig. 4.13b, d) FIB repair. Residual of photoresist is visible at the repaired side due to the Ga staining (Fig. 4.13d). An additional etching is required to remove the stained layer and recover the transparency of glass substrate [49].

With circuit dimension further shrinking, electron beam assisted etching and deposition became popular for IC photomask repair [50], particularly for the EUV mask repair [51]. The focused e-beam is applied in the presence of certain precursor gas molecules, thus triggering the local removal or deposition of materials. With electron beam, not only the repair resolution has increased but also the gallium ion staining issue has been avoided. Because the e-beam assisted repair is basically a

chemical process, it has much better selectivity in removing different materials than the simple physical sputtering by FIB. There were investigations using the FIB systems based on gas field ion source (GFIS) to repair EUV masks, as GFIS-based FIB is also of high resolution to be able to repair defects as small as 11 nm [52], and the helium or neon ions do not create staining issue. However, it was found that the gaseous ion sputtering could cause subsurface damage such as beam-induced mixing of the Mo-Si layers and nanobubble formation in the underlying Mo-Si multilayer of EUV mask [53]. Therefore, e-beam tool has become the only choice for EUV mask repair [54].

4.6.3 Preparing TEM/STEM Samples

Transmission electron microscope (TEM) or scanning transmission electron microscope (STEM) requires very thin samples so that electrons can penetrate the sample to form electron diffraction patterns. Conventional method to prepare TEM samples is mechanical cutting and grinding. Such technique can only prepare large size of samples. With FIB technology, a sample can be "sliced" locally for observation and analysis. The preparation of TEM samples by FIB is similar to cross-sectioning process in IC inspection. Ion beam sputtering is carried out from front and back directions, eventually leaving a thin slice in the middle as the TEM sample. Figure 4.14 demonstrates a prepared TEM sample using a Thermo Fisher's FIB tool [55]. By controlling the precision of sputtering area, the sample thickness can be made of only 100 nm.

Because the direction of ion sputtering is perpendicular to the surface of TEM sample, there are no implanted gallium ions in the sample. However, other damages to the sample can occur, particularly the footprints left by ion beam sputtering parallel to the sample surface. If the ion beam diameter and scanning step size are large, which is often the case when high beam current is required to remove large quantity of material, microscale ripples can be formed along the sputtering direction

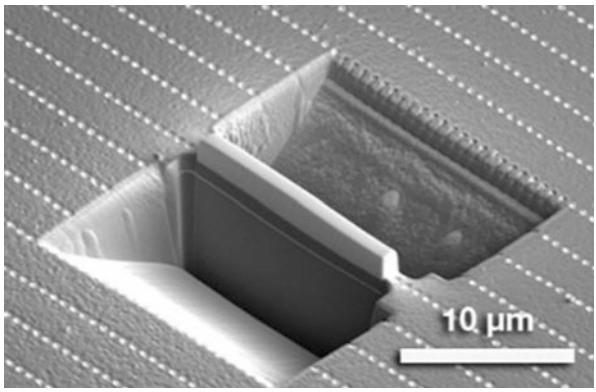


Fig. 4.14 TEM sample prepared by controlled FIB sputtering. (Courtesy of Thermo Fisher Scientific)

which is parallel to the sample surface. When under TEM, the ripple-like surface topography on the sample can degrade the image contrast of actual crystalline structure of the sample, because the periodic nature of the ripples may confuse the periodic crystalline structures of the sample. Such issue is particularly serious for ultrathin samples. Therefore, when making TEM samples with FIB, a low resolution ion beam is used first to sputter in a large area to remove bulk of the material; then a fine ion beam and small scan steps are used to carry out a final polish of the sample surface. In the latest commercial FIB-SEM dual-beam tools, TEM/STEM sample preparation process has been programmed into the systems, and the sample preparation becomes fully automatic [55].

4.6.4 Nanostructuring for Scientific Research

Apart from the aforementioned three applications of FIB, which are the industrial standard processes, scientific research community also encounters the needs for making various micro- and nanostructures. Optical or electron beam lithography introduced in Chaps. 2 and 3 can only create the images of these structures in polymer resists, which have to be transferred to a functional material by etching or deposition. FIB can bypass the pattern transfer step and make micro- and nanostructures directly by either sputtering or deposition; in particular, FIB can make true 3D nanostructures. For example, with FIB-induced deposition, 80 nm pillar with the height of several micrometers can be made, and the pillars can be further manipulated into various shapes and bending angles [56]; materials which are not easy to structure in nanoscale by other techniques, such as diamond, can be milled into 3D shapes by FIB [57]; freestanding structures can be folded by FIB irradiation which alters the internal stress of structure locally [58]; FIB milling helped to break up the single grain boundary of gold film and created a gap as narrow as 1–2 nm [59]. One can find many pictures of exotic structures created by FIB on the Internet. A recent review article has given a comprehensive summary about the variety of applications realized by FIB [60]. With the unique capability of FIB-induced modifications, new nanomaterials with tailored properties can also be created, leading to many more new applications [20].

4.7 Focused Ion Beam Lithography

Like electron beam lithography, focused ion beam can also be used for lithography purpose. Ion beam lithography is of very high sensitivity because energy transfer between ions and solid atoms is much more efficient than between electrons and solid atoms. Commonly used electron beam resists are 100 times more sensitive to ion beam than to electron beam. For example, a resist may need 20 $\mu C/cm^2$ exposure dose by 20 keV electron beam lithography. The same resist only needs 0.1 $\mu C/cm^2$

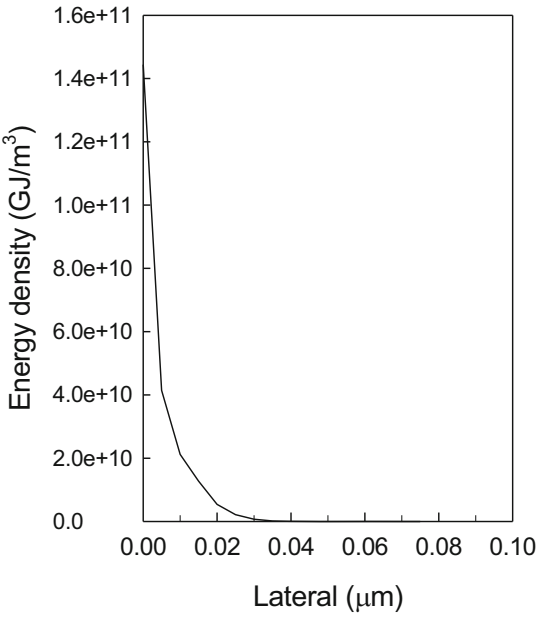


Fig. 4.15 Point spread function of ion scattering (computer simulation result)

exposure dose by gallium ion beam exposure at 100 keV [61]. In addition to high sensitivity, ion beam lithography almost has no proximity effect. Because ions are much heavier than electrons, the scattering range of ions in resist is much less than electrons. There are little spread of ions due to forward scattering and almost no backscattering ions. Figure 4.15 is the computer-simulated point spread function of ion scattering. Comparing to the point spread function of electron scattering (Fig. 3.14 in Chap. 3), the lateral energy spread caused by ion scattering is much less, and there is no "long tail" of energy spread due to backscattering. Experiments have proved that whether for dense line exposure or exposure on substrate of high atomic number material, which is the major cause of proximity effect in e-beam lithography, feature dimensions achieved by ion beam lithography are the same as with light substrate material and for sparse lines [61]. With ion beam lithography, the smallest features of 11–14 nm were demonstrated more than 30 years ago [62].

Focused ion beam lithography with liquid gallium ion sources once drew some interests in its early development stage. It did not prosper because gallium ions have very small penetration depth in resist, as is demonstrated by Monte Carlo simulation in Fig. 4.5. The only way to increase the penetration depth is to either increase the beam energy or to use lighter ions. The industry has gone along the direction of using helium ions generated by gas field ion source (GFIS) instead of gallium LMIS for ion beam-based lithography. The simulation by SRIM program [31] indicates that while the penetration depth of 30 keV gallium ions in PMMA resist is only 22 nm, the same energy of helium ions has penetration depth of 200 nm. Helium ions are inert

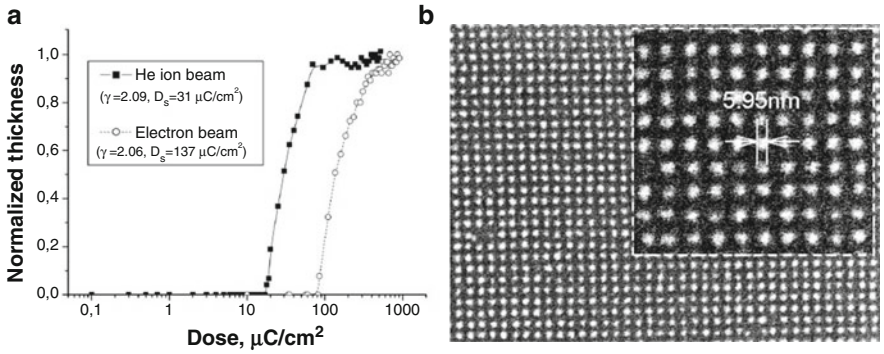


Fig. 4.16 (**a**) Comparison of sensitivity between helium beam and electron beam exposure of HSQ; (**b**) dot array obtained by helium FIB lithography of HSQ

and will not change the property of substrate; therefore, it has no "staining" problem. FIB system with helium GFIS can achieve the smallest beam diameter of 0.25 nm, while FIB with gallium LMIS can only achieve beam diameter around 5 nm. Commercial helium ion microscope equipped with pattern generator from Carl Zeiss AG has been used as helium FIB lithography tool [63].

The advantage of helium FIB lithography has been confirmed by experiment. Figure 4.16a shows the development curves of helium ion beam exposure of HSQ resist in comparison to e-beam exposure. Much less exposure dose is needed when using helium FIB and it has similar contrast as e-beam. It has been shown in Chap. 3 that the highest resolution of e-beam lithography is around 6 nm and is only possible for sparse patterns. Helium FIB lithography can achieve the same level of resolution but on dense patterns, as shown in Fig. 4.16b where dense dots of 5.95 nm were obtained with helium FIB exposure of HSQ resist [64]. A direct comparison of exposure on the same resist (10 nm thickness) with both a helium ion beam (30 keV) and an electron beam (100 keV) showed that the helium FIB could achieve 5 nm dense lines with 20 nm pitch, while the electron beam could only get 6.5 nm dense lines with 40 nm pitch [65].

Although focused ion beam lithography has superior resolution capability, to become mainstream lithography technology, throughput is the key hurdle. Similar to electron projection lithography (EPL), ion projection lithography (IPL) also once received great attention in the late 1990s. IPL was proposed as an optional technology competing with EPL for next generation lithography (NGL) [66]. The main reason to develop IPL was due to the aforementioned facts, i.e., ions scatter little in resist and substrate, hence no proximity effect, and resists are several orders of magnitude more sensitive to ions than to electrons. The first IPL system ALG-1000 using helium ions was made in 1996. A series of development work was carried out in the following years. By 2002, 50 nm pattern resolution was demonstrated by IPL [67]. However, IPL suffers the same drawback as the EPL, which is the need of a

mask. For IPL, a stencil mask has to be used because ions cannot penetrate through a thin membrane.

In order to avoid the bottleneck of mask, multiple beam approach was developed, which was called the Projection Mask-Less Patterning (PMLP) [68]. The PMLP tool can produce up to 40,000 programmable beams working parallel, using H+, He+, Ar +, C60-, or cluster ions. With individual beam diameter of 8.3 nm, the system is able to expose 12.5 nm half-pitch dense patterns on 20-nm-thick HSQ resist [69]. Though of very high resolution and free of proximity effect, it suffered the same fate as multi-e-beam lithography, i.e., not able to compete with multi-patterning optical lithography in terms of throughput. The maturity of EUV lithography finally killed the hope of charged beams going into IC manufacturing. The company (IMS Nanofabrication GmbH) who developed the PMLP tool turned its focus on developing multi-e-beam lithography tool since 2010 [70], and now their multi-beam mask writer (MBMW) becomes the workhorse for making the latest optical and EUV masks, which has been introduced in Sect. 3.7.2 of Chap. 3.

References

1. Swanson, L.W.: Liquid metal ion sources-mechanism and application. Nucl. Instr. Methods Phys. Res. **218**(1–3), 347 (1983)
2. Clampitt, R., Jefferies, D.K.: Miniature ion source for analytical instruments. Nucl. Inst. Methods. **149**, 739 (1978)
3. Swanson, L.W., Schwind, G.A., Bell, A.E.: Emission characteristics of a liquid gallium ion source. Scan. Electron Microsc. **1979/I**, 45 (1979)
4. Kingham, D.R., Swanson, L.W.: A theoretical model of a liquid metal ion source. Vacuum. **34**, 941 (1984)
5. Cui, Z., Tong, L.: A new approach to simulating liquid metal ion sources. J. Vac. Sci. Technol. **B6**(6), 2104 (1988)
6. Cui, Z., Tong, L.: Dynamic characteristics of liquid metal ion sources. J. Vac. Sci. Technol. **B7**(6), 1813 (1989)
7. Bischoff, L., Akhmadaliev, C.A.: An alloy liquid metal ion source for lithium. J. Phys. D. Appl. Phys. **41**, 052001 (2008)
8. Orloff, J., Utlaut, M., Swanson, L.W.: Physics of Liquid Metal Ion Sources, in High Resolution Focused Ion Beams. Kluwer Academic Publisher (2003)
9. Bauerdick, S., et al.: Multispecies focused ion beam lithography system and its applications. J. Vac. Sci. Technol. B. **31**(6), 06F404–06F401 (2013)
10. Sudraud, P., Orloff, J., Benassayag, G.: The effect of carbon bearing gases and secondary electron bombardment on a liquid metal ion source. J. Phys. **47**(11), 381 (1986)
11. Smith, N.S., et al.: Advances in source technology for focused ion beam instruments. MRS Bull. **39**, 329–335 (2014)
12. Smith, N.S., et al.: High brightness inductively coupled plasma source for high current focused ion beam applications. J. Vac. Sci. Technol. **24**(6), 2902–2906 (2006)
13. Kellogg, S.M., et al.: High throughput sample preparation and analysis using an Inductively Coupled Plasma (ICP) focused ion beam source. Microsc. Microanal. **16**(2), 222–223 (2010)
14. Economou, N.P., et al.: The history and development of the helium ion microscope. Scanning. **34**, 83–89 (2012)
15. Ward, B.W., et al.: Helium ion microscope: a new tool for nanoscale microscopy and metrology. J. Vac. Sci. Technol. B. **24**(6), 2871–2874 (2006)

16. Hlawacek, G., et al.: Helium ion microscopy. J. Vac. Sci. Technol. B. **32**(2), 020801 (2014)
17. He, S., et al.: Helium-ion-beam nanofabrication: extreme processes and applications. Int. J. Extrem. Manuf. **3**, 012001 (2021)
18. Livengood, R.H., et al.: The neon gas field ion source – a first characterization of neon nanomachining properties. Nucl. Inst. Methods Phys. Res. A. **645**, 136–140 (2011)
19. Lai, W.-C.: Xenon gas field ion source from a single-atom tip. Nanotechnology. **28**(25), 255301 (2017)
20. Zhao, L., et al.: The 3D controllable fabrication of nanomaterials with FIB-SEM synchronization technology. Nanomaterials. **13**, 1839 (2023)
21. Helios FIB-SEM dual beam systems from Thermo Fisher Scientific Inc. https://www.thermofisher.com/
22. FIB-SEM from Raith. https://raith.com/
23. NX9000 FIB-SEM system from Hitachi. https://www.hitachi-hightech.com/
24. FIB systems from JEOL. https://www.jeol.com/
25. ORION NanoFab from Zeiss. (https://www.zeiss.com/)
26. Lindhardt, J., Scharff, M., Schiott, H.: Range concepts and heavy ion ranges. Mat. Fys. Medd. Dan. Vid. Selsk. **33**(14), 39 (1963)
27. Lindhardt, J., Scharff, M.: Energy dissipation by ions in the keV region. Phys. Rev. **124**, 128 (1961)
28. Cui, Z.: A Monte Carlo Programme for Simulating Ion Penetration in Amorphous Solid Targets. CMF Internal Report (1993)
29. INTERACTIONS OF IONS WITH MATTER – Free downloadable computer simulation program SRIM. http://www.srim.org/
30. Biersack, J.P., Haggmark, L.: A Monte Carlo computer program for the transport of energetic ions in amorphous targets. Nucl. Instr. Methods. **174**, 257 (1980)
31. Zieglera, J.F., Zieglerb, M.D., Biersackc, J.P.: SRIM – the stopping and range of ions in matter (2010). Nucl. Instrum. Methods Phys. Res. B. **268**(11–12), 1818–1823 (2010)
32. FIB Systems for Advanced Circuit Editing. 2002. Available from: FEI Company: http://www.fei.com/
33. Sigmund, P.: Theory of sputtering I: sputtering yield of amorphous and polycrystalline targets. Phys. Rev. **184**, 383 (1969)
34. Orloff, J., Utlaut, M., Swanson, L.W.: Applications of Focused Ion Beams. In: High Resolution Focused Ion Beams. Kluwer Academic Publisher (2003)
35. Utke, I., et al.: Gas-assisted focused electron beam and ion beam processing. J. Vac. Sci. Technol. B. **26**(4), 1197–1246 (2008)
36. Fu, X., et al.: Gas assisted etching of niobium with focused ion beam. Microelectron. Eng. **78–79**, 29–33 (2005)
37. Crow, G.A.: Endpoint detection and microanalysis with Ga FIB SIMS. Proc. ISTFA'91, 401 (1992)
38. Ochiai, Y., Gamo, K., Namba, S.: Characteristics of ion beam assisted etching of GaAs using focused ion beam: dependence on gas pressure. Jpn. J. Appl. Phys. **23**(Pt2, 6), L400 (1984)
39. Santschi, C., et al.: Interdigitated 50 nm Ti electrode arrays fabricated using XeF2 enhanced focused ion beam etching. Nanotechnology. **17**, 2722–2729 (2006)
40. Menon, R., Nabhiraj, P.Y.: High speed micro-fabrication using inductively coupled plasma ion source based focused ion beam system. Vacuum. **111**, 166–169 (2015)
41. Stewward, D., et al.: Focused ion beam gas assisted deposition for mask repair. Proc. SPIE. **1089**, 18 (1989)
42. Blauner, P., et al.: Focused ion beam induced deposition of low resistivity gold films. J. Vac. Sci. Technol. **B7**, 816 (1989)
43. Gandhi, A.: Characterisation of focused ion beam induced processes: etching and metal deposition (PhD Thesis). Oregon Graduate Institute (1991)
44. Tao, T., Ro, J., Melngailis, J.: Focused ion beam deposition of platinum. J. Vac. Sci. Technol. **B8**, 1826 (1990)

45. Prewett, P.D., et al.: Gallium staining in FIB repair of photomasks. Microelectron. Eng. **21**, 191 (1993)
46. Livengood, R., Tan, S., Greenzweig, Y.: Subsurface damage from helium ions as a function of dose, beam energy, and dose rate. J. Vac. Sci. Technol. B. **27**, 3244 (2009)
47. Mohiuddin, T.: FIB circuit edit becomes increasingly valuable in advanced node design. Electronic Design (Jan 2014)
48. Akerson, D.: Circuit editing evolves to meet semiconductor needs. In: Wong, W.G. (ed.) Electronic Design (2022)
49. Cui, Z., Prewett, P.D., Watson, J.G.: Focused ion beam biased repair of conventional and phase shift masks. J. Vac. Sci. Technol. **14**(6), 3942 (1996)
50. Klaus, E., Hans, B., Rainer, B.: A novel electron-beam-based photomask repair tool. Proc. SPIE. **5256**(1222) (2003). https://doi.org/10.1117/12.532866
51. Waiblinger, M.: E-beam induced EUV photomask repair – a perfect match. BACUS News. **26**(5) (2010)
52. Aramaki, F., Kozakai, T., Matsuda, O.: Photomask repair technology by using gas field ion source. BACUS News. **29**(14) (2013)
53. Gonzalez, C.M., et al.: Evaluation of mask repair strategies via focused electron, helium, and neon beam induced processing for EUV applications. Proc. SPIE. **9048** (2014). https://doi.org/10.1117/12.2046712
54. Heil, T., et al.: Pushing the limits of EUV mask repair: addressing sub-10 nm defects with the next generation e-beam-based mask repair tool. J. Micro/Nanopattern. Mater. Metrol. **20**(3), 031013 (2021)
55. Sample preparation for TEM and STEM imaging of semiconductor devices. https://www.thermofisher.com/
56. Li, W., Changzhi, G., Cui, A.: Three-dimensional nanostructures by focused ion beam techniques: fabrication and characterization. J. Mater. Res. **28**(22), 3063 (2013)
57. Jiang, Q., Liu, D., Liu, G.: Focused-ion-beam overlay-patterning of three-dimensional diamond structures for advanced single-photon properties. J. Appl. Phys. **116**, 044308 (2014)
58. Cui, A., Liu, Z., Li, J.: Directly patterned substrate-free plasmonic nanograter structures with unusual Fano resonances. Light Sci. Appl. **4** (2015). https://doi.org/10.1038/lsa.2015.81
59. Cui, A., Liu, Z., Dong, H.: Single grain boundary break junction for suspended nanogap electrodes with gap width down to 1–2 nm by focused ion beam milling. Adv. Mater. **27**, 3002 (2015)
60. Li, P., et al.: Recent advances in focused ion beam nanofabrication for nanostructures and devices: fundamentals and applications. Nanoscale. **13**, 1529 (2021)
61. Matsui, S., et al.: Lithography approach for 100 nm fabrication by focused ion beam. J. Vac. Sci. Technol. **B4**, 845 (1986)
62. Gamo, K.: Focused ion beam lithography. Nucl. Instrum. Meth. **B65**, 40 (1992)
63. Maas, D., van Veldhovena, E., Chenb, P.: Nanofabrication with a helium ion microscope. Proc. SPIE. **7638**, 763814-1 (2010)
64. Sidorkin, V., van Veldhoven, E., van der Drift, E.: Sub-10-nm nanolithography with a scanning helium beam. J. Vac. Sci. Technol. B. **27**(4), L18 (2009)
65. Cattoni, A., et al.: Sub-10 nm electron and helium ion beam lithography using a recently developed alumina resist. Microelectron. Eng. **193**, 18–22 (2018)
66. Gross, G.: Ion projection lithography: next generation technology? J. Vac. Sci. Technol. **B15**(6), 2136 (1997)
67. Hirschera, S., et al.: Ion projection lithography below 70 nm: tool performance and resist process. Microelectron. Eng. **61–62**, 301–307 (2002)

68. Platzgummer, E., Löschner, H., Gross, G.: Projection maskless patterning (PMLP) for the fabrication of leading-edge complex masks and nano-imprint templates. Proc. SPIE. **6730**, 108 (2007)
69. Platzgummer, E.: Maskless lithography and nanopatterning with electron and ion multi-beam projection. Proc. SPIE. **7637** (2010)
70. Multi-Beam Mask Writer (MBMW). https://www.ims.co.at/

Chapter 5
Tip-Based Nanofabrication



5.1 Introduction

As has been described in Chaps. 2, 3, and 4, photons and charged beams (electron and ion) are capable of delineating sub-100 nm patterns. However, one question has been missing, that is, at what cost? To break into sub-100 nm scale, many "tricks," apart from short wavelength and high NA, have to be used in DUV-based lithography, and the latest high-NA EUV lithography tool costs over 350 million euros. In charged beam (electron or ion) lithography, very sophisticated electron or ion optical systems are required; so is the escalating cost to own such a system. For low-cost nanoscale patterning technologies and throughput is not a critical issue, scanning probe lithography (SPL) is definitely an alternative and accessible to majority of scientific research community.

The advent of SPL was shortly after the invention of scanning tunneling microscope (STM). Researchers actually stumbled onto it when they observed that some lines were left on the sample surface after an extended period of scanning of STM probe [1]. It is now known as local oxidation process initiated by a probe. At the time when the lines were observed, the mechanism was unclear. It was said "under the influence of air some oxidation process seems to occur and maybe SiO_2 is formed." Although the SiO_2 lines drawn by the STM tip were nowhere comparable in quality to those patterned by optical or e-beam lithography, it was the start of drawing patterns at sub-100 nm scale without photons or charged beams, and a STM was a very simple and cheap setup. The invention of atomic force microscope (AFM) further fueled the development. In the last 40 years since the advent of STM, many new nanofabrication techniques have been developed using a scanning tip which are no longer necessary to be an imaging tool such as STM or AFM. It is now simply called tip-based nanofabrication (TBN) [2].

Conventional lithography using either photons or electrons always has to rely on a polymer material (resist) as an imaging layer. TBN, however, can be implemented with diverse mechanisms [3]. Thin polymer resist can be exposed by electrons

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_5

emitted from a tip, light transmitted through a probe or simply by heating; the surface of a sample can be modified under the influence of very high local electric field, such as local oxidation and depassivation; materials can be transported from a probe to the sample surface to form deposition patterns; patterns are formed simply by scratching the surface mechanically or by pushing surface atoms around with a tip. Though some of the patterning techniques have been commercially exploited as dedicated nanolithography tools, most of the techniques are still confined to laboratories as an add-on function of a scanning probe microscope (SPM).

The most attractive feature of TBN is its simplicity. It makes the sub-100 nm patterning accessible to anyone with a simple setup of scanning tips. However, one has to be aware of its limitation. Firstly, tip-based lithography does not have the same resolution capability as the atomic resolution imaging capability of a SPM, because the mechanism of sensing the existence of an atom is different from making the structural change of surface. Secondly, the very proximal interaction between a tip and a sample surface has determined that the interaction is extremely "superficial." Patterns made by TBN are never more than a few tens of nanometers deep. Therefore, special measures have to be taken when transferring the "superficial" patterns into a substrate. Thirdly, imaging with a SPM is slow; so does the TBN. In fact nanofabrication using a tip is even slower because there has to be enough energy transfer from the tip into a substrate in order to induce a structural change of the surface. TBN generally has the scanning speed of micrometer to millimeter per second, which is many orders of magnitude slower than a conventional e-beam lithography. Fourthly, the working area of a TBN is small, making it suitable only for patterning a small area of surface. There have been continuous efforts to develop multiple probe systems [4] or high-speed scanning systems [5] in order to boost the throughput of TBN, but it can never be successful as nanofabrication tools for industrial-scale applications. Nevertheless, TBN is quite handy for making sub-100 nm structures in laboratories.

5.2 Principles of Scanning Probe Microscopes

Scanning tunneling microscope (STM) was invented in 1981 by Binnig and Rohrer [6], who received Noble Prize in Physics for the invention in 1986. Figure 5.1 schematically illustrates the basic configuration of a STM. It consists of a sharp conductive tip (stylus) which is usually a tungsten wire sharpened electrochemically. The tip assembly is controlled by a piezoelectric transducer which moves the tip up and down to control the gap between the tip apex and the sample surface with sub-nanometer precision. Although the radius of tip apex may be a few nanometers or a few tens of nanometers, at the very end of the tip, there are always a few atoms of tip material protruding closest to the sample surface at a distance about 0.5 to 1 nm, as illustrated by the inset picture in Fig. 5.1. When a bias voltage (mV to a few volts) is applied between the tip and the sample surface, a tunneling of electrons between the tip atoms and sample surface atoms can occur if their gap distance is within

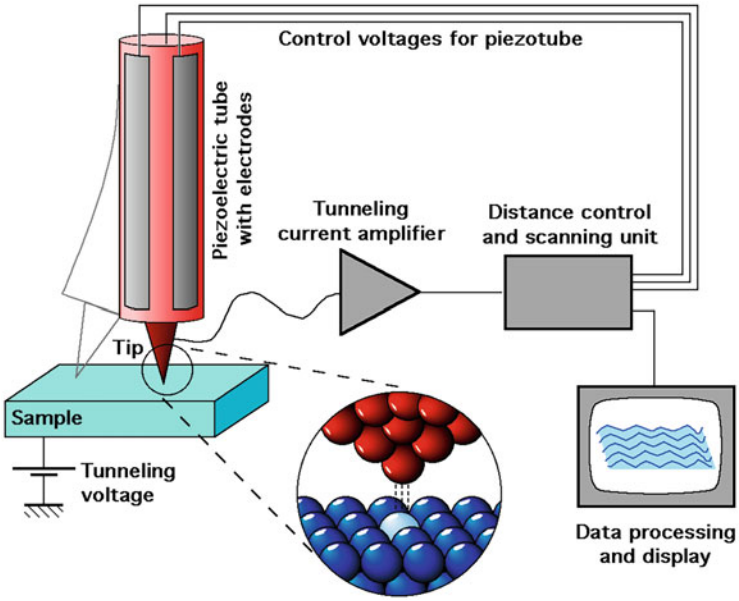


Fig. 5.1 Schematic view of STM. (Reprint from Wikipedia)

~1 nm, as explained by quantum mechanics principle. For suitably sharpened tip, gap distance, and bias voltage, the electron tunneling can be confined only between single atoms. An additional piezoelectric transducer in the tip assembly can produce raster scan of the tip over small region of the sample surface, creating a 2D image of the sample surface with atomic resolution.

The key technique in a STM operation is the feedback control of tunneling current. As the tunneling current is exponentially dependent on the tip-sample distance, the feedback control loop can either keep a constant tunneling current or keep a constant tip-sample distance, with sub-angstrom precision vertically and sub-nanometer precision laterally. In the constant current mode, tip-sample distance across the scanning region is varied, creating a 2D topographical map of the surface. In the constant gap mode, variation of electron density of the surface atoms is detected. In either case, a 2D map of surface atoms can be created.

Atomic force microscope (AFM) was invented shortly after the STM [7]. As schematically illustrated in Fig. 5.2, the key difference between a STM and a AFM is the tip assembly. In an AFM, the tip is mounted on a flexible cantilever beam, whereas the STM tip is rigidly mounted. AFM operates by measuring attractive or repulsive forces between the tip and the sample surface. The forces cause the cantilever beam to deflect, which is detected optically or electronically and translated to surface topographical variation. A raster scan of the tip over a surface region generates a 2D map of the surface structure with atomic resolution. AFM can be operated either in contact mode or noncontact mode which is also called tapping mode. In contact mode, a feedback control loop can either keep a constant height of

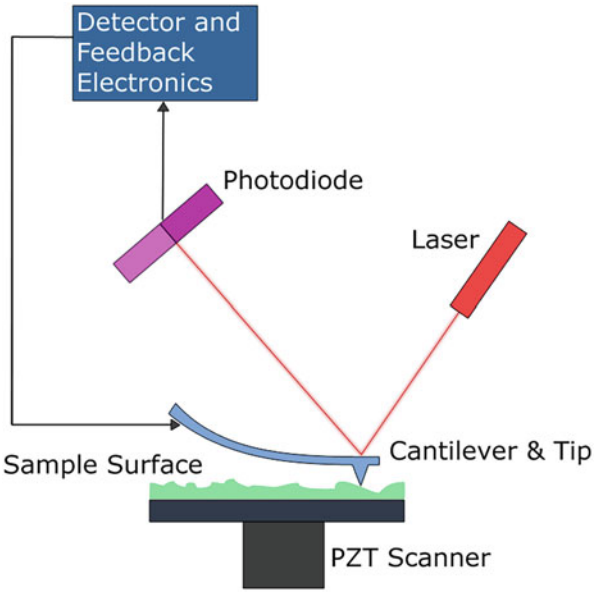


Fig. 5.2 Schematic view of AFM. (Reprint from Wikipedia)

the tip over the sample surface or keep a constant contact force between the tip and sample surface. In the tapping mode, the AFM tip is literally tapping the sample surface at high frequency. In all these modes, a 2D map of surface topographical image can be created.

AFM is actually a much more versatile technique than the STM. A STM relies on tunneling electrons for imaging, which requires a vacuum environment. AFM can be operated in air. STM requires both the tip and sample are conductive, while AFM can be used to image any materials, even in liquid. With a conductive tip, AFM can be operated like a STM as well, to provide typical capabilities of a STM. Over the years, many other types of scanning probe microscopic techniques have evolved from the basic concept of AFM, such as measuring lateral force using a cantilever tip, magnetic force, electrostatic force, Kelvin potential, surface conductivity, capacitance, temperature, etc. [8].

In addition to STM and AFM, optical fiber tip can be used to transmit light, which is the near-field scanning optical microscope (NSOM), or sometimes called scanning near-field optical microscope (SNOM). Invented also shortly after the advent of STM [9], NSOM uses a sharpened light source as a scanning probe which is positioned only a few nanometers above the imaging surface. The light source is usually a laser beam transmitted through optical fiber to the tip aperture. A constant height of fiber probe above a sample surface is maintained by a shear force feedback system which employs a tuning fork commonly found in digital watches. The oscillation of tuning fork is damped as a result of contact shear force, and this effect is detected and fed back into the control electronics to adjust the height of the probe. NSOM can also be realized by drilling a hole on an AFM tip and shine laser light

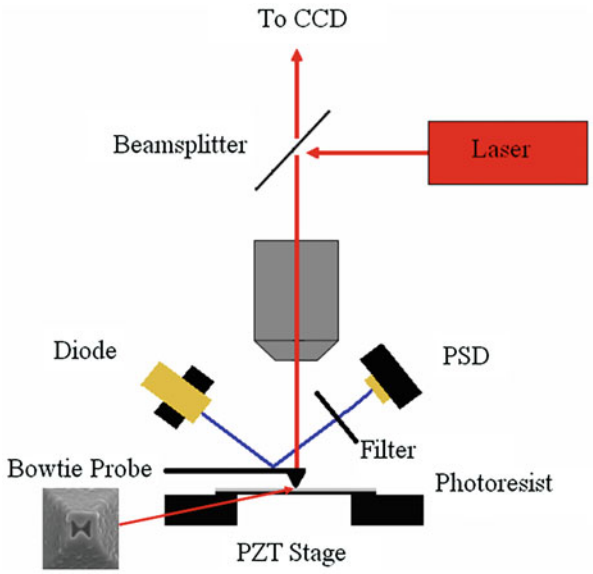


Fig. 5.3 Cantilever-type NSOM with a plasmonic bowtie probe

through it, which then operates the same as an AFM, as schematically shown in Fig. 5.3.

It should be noted that the optical fiber tip must be coated with metal (usually aluminum but sometimes other metals such as gold) to prevent leakage of light laterally through the tip sidewall and allows only a small opening aperture of just a few tens of nanometers at the end of glass fiber to let the laser light out. However, the extremely small aperture can severely restrict the light output, which limits the energy delivered to the surface, hence the effectiveness of nanofabrication. Using plasmonic field enhancement produced from a bowtie aperture (Fig. 5.3) can generate peak intensities orders of magnitude larger than that found from circular apertures with similar optical confinement [10].

Although all the three SPMs have been explored for nanofabrication in the past, nowadays majority of reported nanofabrications were done by AFM type of scanning probes. The cantilever-based tips are mechanically flexible and can work either as a STM or a NSOM. The progress in silicon micromachining processes allows to make the cantilever and tips easily. In particular, a variety of functions can be integrated onto the cantilever, such as piezo- or magnetic stimulation of bending, electronic detection of bending, and heating of the tips. Many types of TBN have been reported, including exposure of resists by electrons, photons, and even heat; modification of surface electrically or atomically; transporting materials from the tip to substrate; or just scratching the surface. Although the TBN has high resolution capability, normally below 100 nm, and accurate alignment as well as positioning because they are microscopes, they all suffer the drawback of low throughput and limited pattern area. To overcome the drawback, considerable efforts have been

devoted to developing massively parallel scanning tips to boost the throughput in the past decades. In the following sections, four types of TBN are introduced, together with some high-throughput techniques.

5.3 Exposure of Resists

5.3.1 Exposure by Electrons

It is well known that electron emission can occur at very high electrical field, which is called field emission. The SEMs and e-beam lithography systems have all been utilizing field emission cathodes to produce high current density and low energy spread electrons, which have been described in Chap. 3. Field emission is governed by the Fowler-Nordheim theory [11], as shown in Eq. (5.1).

$$J = k_1 E^2 \exp\left(\frac{-k_2}{E}\right) \tag{5.1}$$

where J is the field emission current density, E is the field strength, and k_1 and k_2 are constants related to the work function of field emitter material. There is a threshold field (~2×10^7 V/cm) above which electron emission starts. With the increase of field strength, the emission current density increases rapidly. There are two ways to achieve high field strength on an electron emitter: using a very sharp emitter or brining the counter electrode as close as possible to the emitter. Figure 5.4a shows an example of emitter and counter electrode configuration with the inset SEM picture of an actual silicon tip (scale bar 500 nm). The emitter has a radius of 50 nm at its apex and the counter electrode is above the emitter with a few tens of nanometers gap. The

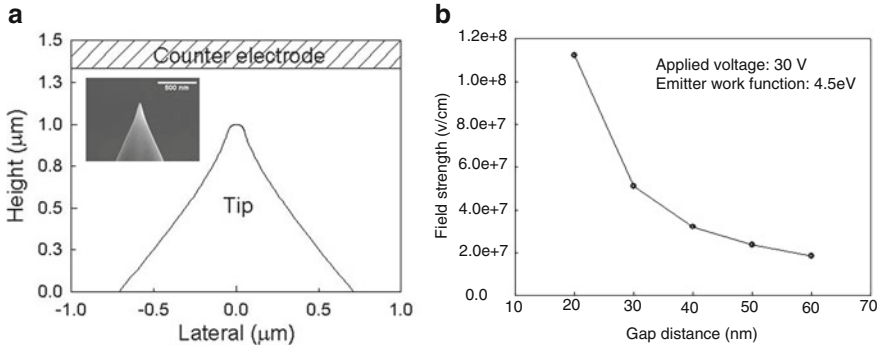


Fig. 5.4 (**a**) Schematic of field emitter and counter electrode with the inset SEM picture of an actual silicon tip (scale bar 500 nm); (**b**) field strength vs. gap distance between the tip and counter electrode

electrical field at the surface of emitter is numerically calculated using a home-developed computer simulation program [12], based on the geometry shown in Fig. 5.4a. Figure 5.4b shows the surface field strength as the function of gap distance between the emitter and the counter electrode. It is apparent that the field strength can be above the field emission threshold (~2×10^7 V/cm) even at 30 V of applied voltage, as long as the emitter tip is close to the counter electrode within 50 nm range. In the meantime, the simulation also showed that the sharper the tip, the higher the field strength at a fixed counter electrode distance [12]. This establishes the basis for electron exposure of resists using a tip.

Electron exposure with a sharp tip is simply to move the tip on top of a conductive substrate coated with a thin layer of electron-sensitive polymer, while a bias voltage is applied between the substrate and the tip [13]. Exposure of PMMA resist with a STM was first reported in the late 1980s. Lines of 22 nm wide were made with exposure of 20-nm-thick PMMA resist and scanning speed of 1 μm/s [14]. Exposure of chemically amplified resist SAL601 using a STM was also reported [13]. When exposing a resist, the STM is normally operated at constant current mode. The applied bias voltage to the tip is much higher, usually a few tens of volts, than it is operated as a microscope, which is normally a few volts.

Exposure of resists with a field emission tip is no different from a conventional e-beam lithography described in Chap. 3, except that electrons emitted from a tip are of very low energy and in close proximity to the resist layer. There are some unique features one has to be aware in tip-based e-beam lithography.

1. Very thin resist layer (<50 nm) has to be used. As seen in Fig. 5.4a, the tip has to be very close to the substrate which acts as the counter electrode, in order to generate field electron emission at low bias voltage. The tip and the sample substrate are only separated by the resist layer which has to be very thin. On the other hand, the emitted electrons have kinetic energy of only a few tens of volts. They cannot travel very far in resist, as explained in Chap. 3, which requires the resist layer to be thin so that it can be fully exposed.
2. Unlike conventional e-beam system where a series of electron lenses are employed to focus the electrons into a very fine beam, a STM does not have lenses between its tip and substrate. Electrons emitted from the STM tip are generally divergent. The degree of divergence depends on both the applied voltage and the separation between tip and substrate [15]. For example, a tip of 10 nm apex radius forms about a 10-nm-diameter beam if the separation is 10 nm. The beam diameter becomes 30 nm if the separation is 30 nm, which almost linearly increases with the separation. High bias voltage also causes large divergence of emitted electrons, due to the increase of tip-substrate separation in order to maintain constant current working mode. A 10 nm beam at 20 V of bias voltage becomes 40 nm at 40 V bias voltage [15]. Because there has to be a separation between the tip and substrate in order to accommodate the resist layer and the applied bias voltage has to be high enough to generate field electron emission, the resultant beam of electrons is always bigger than the source size. This is the

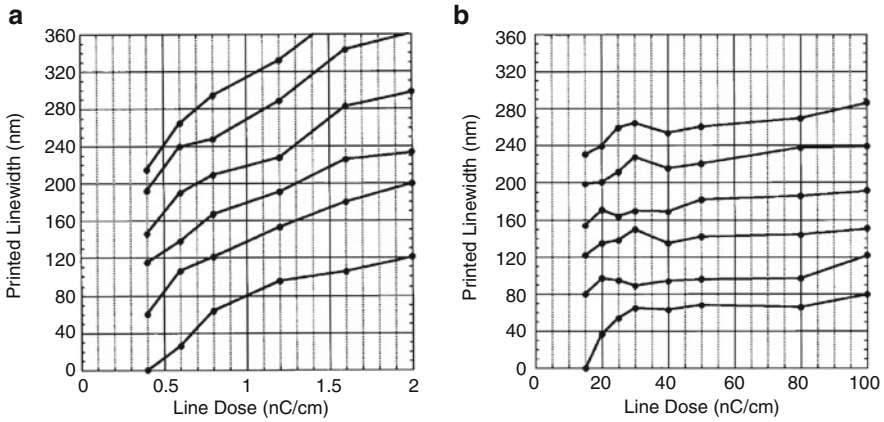


Fig. 5.5 Comparison of line patterns vs. exposure dose (**a**) for conventional e-beam lithography at 30 kV and (**b**) for STM lithography at 40–60 V. (Reprint from [16] with permission)

reason why the resolution of STM lithography can never match the atomic resolution of STM imaging.

3. Low energy exposure is the key feature of tip-based lithography. In conventional e-beam lithography, low energy electrons (<5 keV) can increase the efficiency of resist exposure, or resists appear to be more sensitive to low energy electrons. However, this is not the case in STM exposure of resists. Although the field emission electrons from a STM are normally below 50 eV, the efficiency of resist exposure by such low energy electrons is much lower, or STM exposure of resists requires much higher exposure dose than conventional e-beam lithography. Figure 5.5 compares the linewidth vs. exposure dose by conventional e-beam lithography (Fig. 5.5a) and by STM exposure (Fig. 5.5b). Different linewidth was created by multiple scans of beam in 40 nm grid separation. From the comparison it becomes apparent that STM requires 30 times higher exposure dose than conventional e-beam lithography to expose the same patterns. This implies a fundamentally different exposure mechanism in play with STM exposure of resists. Conventional e-beam lithography systems deliver an e-beam with high energy into a resist layer. The electrons undergo scattering and slow down until they lose all energy. The STM tip, however, emits electrons which have very low initial energy (a fraction of eV) when they enter the resist layer because there is not enough travel distance for the electrons to acquire acceleration energy. They are accelerated when entering the resist layer toward conductive substrate. During the process they lose energy to inelastic scattering with resist molecules as well as gain energy from the high field. Such process is perceived less efficient in breaking the molecular chain of a polymer resist than in the case of electrons with initial high energy. Extremely high exposure dose can only mean slow exposure. Coupled with slow scan speed (~1 μm/s), resist exposure by STM is of very low throughput.

The biggest advantage of low energy e-beam lithography is its low proximity effect or it being free of proximity effect, as already described in Chap. 3. The comparison in Fig. 5.5a, b reveals that STM exposure of resist has much less proximity effect. This is seen from the variation of linewidth with the increase of exposure doses. In conventional e-beam lithography, increase of exposure dose will result in increase of exposed feature dimension because at higher exposure dose, large range scattering electrons acquire enough energy to contribute to the exposure of resists, as seen in Fig. 5.5a. This is not the case for STM exposure of resists. Increase of exposure dose shows nearly no change of linewidth, as seen in Fig. 5.5b, which means no electrons are found outside the exposure region.

Tip-based e-beam lithography is no doubt capable of achieving very high patterning resolution. The resolution is dependent on the tip radius and the resist material, together with the applied voltage between tip and sample and the exposure dose. Linewidths ranging from sub-5 nm to >100 nm have been obtained as well as dense lines with half pitch as narrow as 7 nm and dot sizes of sub-10 nm [17]. A range of polymer resists have been investigated, for example, the most widely used high resolution e-beam resist PMMA. It was found that PMMA can be exposed with field emission tip as normal positive tone pattern at line dose of 200 nC/cm but becomes negative tone at 400 nC/cm or even self-developed at >500 nC/cm [18]. As for the low throughput of tip-based e-beam lithography, this would not be a serious problem for laboratory research in which only a small pattern area and a handful devices are needed. There are solutions for high throughput, which will be introduced later in this chapter.

The thin resist layer may have the difficulty of subsequent pattern transfer by reactive ion etching. Multilayer resist system obviously does not work here, because it increases the gap distance between the STM tip and the conductive substrate. In this case, resists with high etch selectivity can be used. For example, SAL601 resist has higher etch selectivity when used as mask to etch silicon (> 5:1) than PMMA (~1:1). Figure 5.6a shows a 50 nm line etched into 300 nm depth of silicon using SAL601 resist as the etch mask. The calixarene-based resists were also used for STM exposure, which have four times more endurance than PMMA for etching silicon [19]. When PMMA resist is exposed with STM, the only possible route of pattern transfer is by metal lift-off. The lift-off metal pattern can be used as the etch mask to further transfer into silicon. Figure 5.6b is such an example that PMMA was first exposed by a STM and metal line pattern was produced by lift-off. The metal line was then used as a mask to etch into silicon. The pattern transfer by etching or lift-off will be introduced in later chapters.

5.3.2 Exposure by Photons

It is quite natural to think of using an optical probe to deliver photons to realize optical lithography. While near-field scanning optical microscope (NSOM) was invented in the mid-1980s [9, 20], its real application in exposure of photoresist

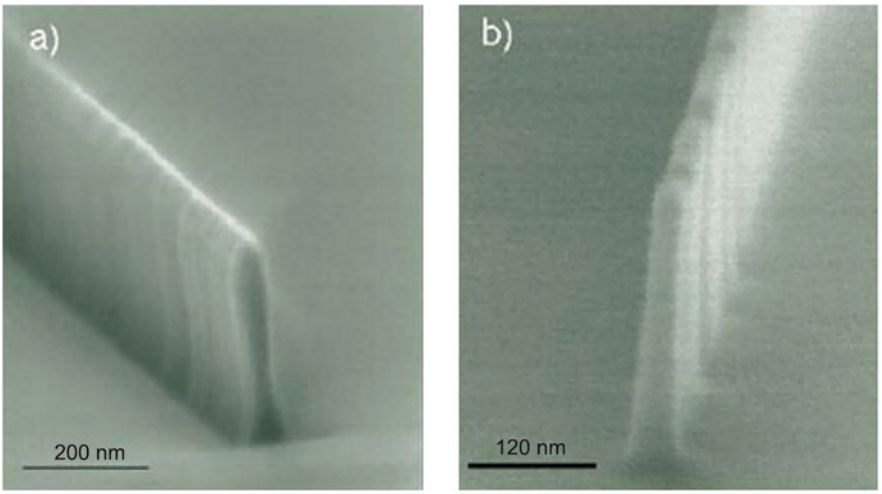


Fig. 5.6 Silicon line patterns produced by STM exposure of SAL601 resist and reactive ion etching (**a**) and PMMA resist with subsequent lift-off of metal line and reactive ion etching using the metal line as mask (**b**). (Reprint from [16] with permission)

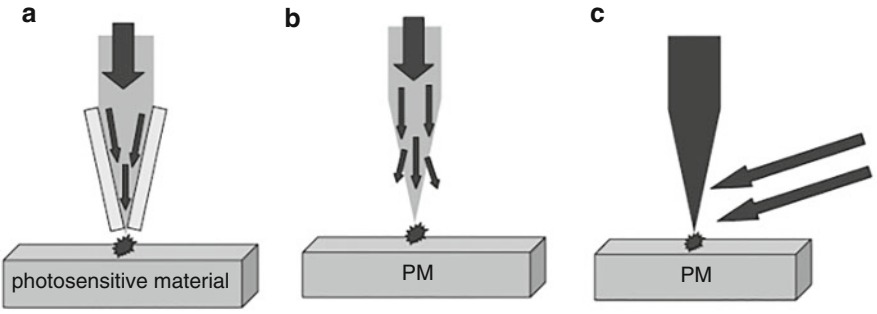


Fig. 5.7 Optical probe types in NSOM: (**a**) glass tip with aperture, (**b**) glass tip without aperture, (**c**) metal tip without aperture. (Reprint from [26] with permission)

was quite a few years later [21]. The exposure is based on near-field evanescent photons, which has been extensively described in Sect. 2.8 of Chap. 2. There are two types of optical probes used for lithography, apertured and apertureless. The NSOM for imaging uses an apertured tip with the aperture size of 50–150 nm. The aperture is created by coating a metal thin film on the surface of a sharpened glass fiber, so that light is only allowed to transmit through the end of fiber, as illustrated in Fig. 5.7a. With thin resist layer, the dimension of exposure feature is only limited by the aperture size of optical probe and the distance between the probe and resist surface. However, it is difficult to make the aperture size smaller than 50 nm. On the

other hand, the light transmission drops rapidly as the aperture gets smaller. Increasing the laser intensity can damage the fiber. To solve the problem, the aperture can be made into a bowtie shape as shown in Fig. 5.3, which can be done by FIB milling [22]. The bowtie aperture that incorporates a small lateral gap between two metal edges can have both tight optical confinement and high optical transmittance enhanced by plasmonic antenna modes (see Sect. 2.8 of Chap. 2). An overall dimension of 160 nm bowtie aperture generated exposed resist linewidth of only 59 nm [10].

An alternative is to use apertureless tip such as the one shown Fig. 5.7b, which is a sharpened optical fiber without metal coating. However, this type of optical probe is less favorable because of the less optimum light confinement in the probe, unless it works in internal reflection mode [23]. The third way is that the probe is simply a metal tip which is illuminated by a light wave, as shown in Fig. 5.7c. Strong localized electrical field can be generated at the extremity of metal tip by the excitation of illuminating light wave, dubbed as "lightning rod" effect [24], which can act as an optical probe. Metal tip is much easier than a glass fiber tip to be made extremely sharp (5–10 nm). A 10 nm resolution lithography has been achieved with metal tip apertureless NSOM [25].

5.3.3 *Exposure by Heat*

As mentioned previously, silicon micromachining allows variety of functions to be built onto a cantilever-type scanning probe. One of the functions is heating. Figure 5.8 shows a cantilever tip with an integrated heater [2]. The cantilever is U-shaped and made of differentially doped silicon such that the cantilever legs are high doped to carry current and the cantilever free end is low doped to form a resistive heater. Passing current through the cantilever causes more than 90% of the power to be locally dissipated at the heater tip, and the heater temperature can reach

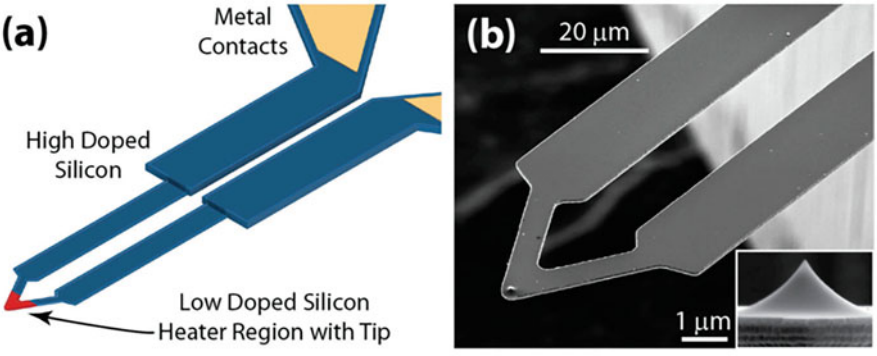


Fig. 5.8 (**a**) Schematic of a cantilever with integrated tip and heater element; (**b**) SEM image of the silicon micromachined cantilever and the tip at the free end (inset)

over 1000 °C. With such highly concentrated heat at the tip apex, many phenomena can happen by tip interacting with substrate surface, including removal of material by thermomechanical indentation or sublimation; conversion of a sample by local modification of its physical properties such as the crystallinity or chemical reaction; and addition of a functional material by melt transfer from a heated tip to the substrate or from a gas phase such as chemical vapor deposition of a precursor material [27].

Analogue to electron or photon exposure of resists, extreme heat can remove polymer materials by direct sublimation without wet development. Heat exposure of resists becomes a major thermal scanning probe lithography (t-SPL) technique. However, conventional resists used in electron or optical lithography such as PMMA require excessively high tip temperatures or long exposure times and begin to flow when heated above the glass transition, which limits their application range. Thermal probe becomes a viable lithography process because of the self-amplified depolymerization polymer: polyphthalaldehyde (PPA) [28]. The molecular structure of PPA is such that upon reaching the glass transition temperature (~150 °C), the polymer chains break and immediately unzip into small volatile molecules, the so-called self-amplified depolymerization. Consequently, PPA does not melt and flow, but rather directly sublimates. The reaction is efficiently activated within a time frame of several μs when using a heater temperature of 700 °C. In addition, the reaction is endothermic, which keeps the depolymerization localized. The patterning depth is a linear function of the applied load, and a mere increase in the force of 3.5 nN leads to an increase in the patterning depth from 0 to 50 nm [29]. The patterned PPA can act as a resist mask to etch into SiO_2 with the etch selectivity of 1:3.3 to SiO_2 [30]. Once the PPA pattern is transferred into SiO_2, the SiO_2 mask then allows further reactive ion etching into silicon or into PMMA for metal lift-off. With this process silicon and metallic structures with sub-20 nm resolution have been produced [31].

5.4 Surface Modification

5.4.1 Hydrogen Depassivation

More than 30 years ago, two IBM researchers Eigler and Schweizer used the ultrasharp tip of a STM to push 35 xenon atoms on a cold nickel surface to form the letters “IBM” [32]. Since then, the atomic precision of scanning probe has been widely recognized and used for manipulation of surface atoms. One of its successful applications is to move hydrogen atoms on silicon surface for lithography purpose.

A fresh surface of silicon on the (100) crystalline plane always has dangling bonds of Si atoms ready to accept foreign atoms to form silicon compounds. That is why there is always a 1.5-nm-thin layer of SiO_2, called native oxide, on the silicon surface once it is exposed to air. Such chemical activity can be passivated by covering the surface with hydrogen atoms, as shown in Fig. 5.9a. Then by selectively

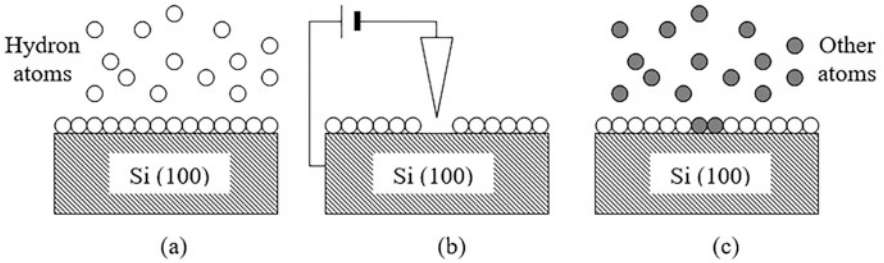


Fig. 5.9 (**a**) Hydrogen passivation of Si (100); (**b**) depassivation by scanning probe; (**c**) absorption of other atoms on the depassivated sites

removing some of the H-atoms by a biased probe to expose the active Si dangling bonds (Fig. 5.9b), atoms of other materials can be introduced to react with the Si atoms to form a compound with specific functions (Fig. 5.9c). This is the so-called hydrogen depassivation lithography (HDL) and the passivation layer of H-atoms acts as the resist [2]. At the sites of exposed Si dangling bonds, atoms of other materials can be deposited by absorption. For example, atomic layer deposition (ALD) of TiO_2 was used to produce an etch-resistant hard mask. Reactive ion etching then transfers the mask pattern onto Si with pattern height of 17 nm, critical dimension of approximately 6 nm, and full pitch down to 13 nm [33]. Since its advent, HDL has been employed by many researchers to fabricate next generation nano-devices such as single-electron transistors, quantum computer qubits, editable atomic-scale memories, and two-dimensional quantum metamaterials [34].

Although HDL can have atomic precision, it needs ultrahigh vacuum to maintain uncontaminated surface and exposure is very slow. As shown in Fig. 5.9b, the depassivation is realized by electrons transferred from the tip to the H-atoms at an applied voltage. There are two distinct desorption mechanisms: (i) direct electronic excitation of the Si-H bond by field-emitted electrons and (ii) an atomic resolution mechanism that involves multiple-vibrational excitation by tunneling electrons at low applied voltages [35]. In the tunneling mode, the tip-sample bias voltage is typically 4 V and an electron current is 4 nA, which generate a line dose of 4mC/cm at the scanning speed of 10 nm/s. By increasing the bias voltage to 8 V and moving the tip away from sample surface (5 nm instead of 1 nm), electrons from the tip change from tunneling mode to field emission mode. The scanning speed can increase to 50 nm/s, albeit the patterning resolution also deteriorates [36].

5.4.2 Field-Induced Oxidation

As described in the previous section, hydrogen-passivated silicon surface can be depassivated with a scanning probe under the influence of tip-emitted electrons. The exposed silicon dangling bonds are then ready to accept foreign atoms. Under

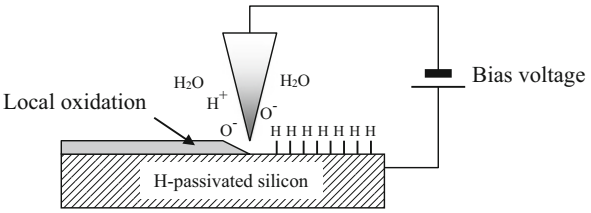


Fig. 5.10 Schematic illustration of field-induced local oxidation

ultrahigh vacuum condition, the foreign atoms are introduced deliberately as is the case of HDL. Under ambient condition, there are inevitably water molecules because of humidity. When the tip is brought to a few nanometers above the surface of a sample, even a few volts of applied voltage between the tip and sample surface can generate electric field up to the magnitude of $\sim 10^8$ V/cm. At such a high electrical field, water molecules can be broken up into ions to become reactive and form silicon dioxide, as schematically illustrated in Fig. 5.10. The local oxidation reaction takes the following steps [37], as expressed in Eq. (5.2).

$$\mathrm{Si} + 2h^+ + 2(\mathrm{OH}^-) \rightarrow \mathrm{Si(OH)_2} \rightarrow \mathrm{SiO_2} + 2\mathrm{H}^+ + 2e^- \tag{5.2}$$

Water molecules (H_2O) are ionized at the high electrical field. Negative ions (OH^-) are driven toward the sample surface (anode). Silicon is oxidized. Positive hydrogen ions are attracted to the tip (cathode) to combine with the released free electrons to become hydrogen gas molecules (H_2).

Local oxidation was first discovered with STM scanning over silicon surface [38]. However, STM is not the best instrument for local oxidation lithography. This is due to the way STM operates. As described earlier, a STM operates either in constant current mode or constant gap mode. The bias voltage control depends on the feedback of either the tunneling current or the tip-sample distance. When oxide grows underneath the tip, both the tunneling current and the tip-sample distance (formation of oxide bump) are changing, which will make the bias voltage change because of the feedback control loop. Damage of STM tip often occurs because the STM often fails to recognize the increase of surface height as the oxide layer grows. When measuring the oxide height during oxidation process, STM can give false measurement, overestimating the height because of the decrease of tunneling current through the oxide layer.

AFM is, on the other hand, better suited for local oxidation lithography. AFM holds two distinct advantages over the STM: (1) the bias voltage which is the root cause of oxidation can be applied independent of the feedback control of the tip-sample spacing; and (2) real oxide height is directly measured without involving tunneling current [39]. AFM working in tapping mode is even better because it eliminates lateral shear forces and overcomes tip-sample adhesion forces (e.g., capillarity). These properties avoid damages on the surface and improve imaging and lithography resolution. Table 5.1 compares the performances of STM, AFM in

Table 5.1 Comparison of STM and AFM for local oxidation lithography

	STM	CM-AFM	TM-AFM
Speed	<8 μm/s	<128 μm/s	<20 μm/s
Resolution	≥10 nm	≥30 nm	≥10 nm
Thickness	≤1.5 nm	≤8 nm	≤4 nm
Reproducibility	<50%	>50%	>75%

Reproduced from [39] with permission

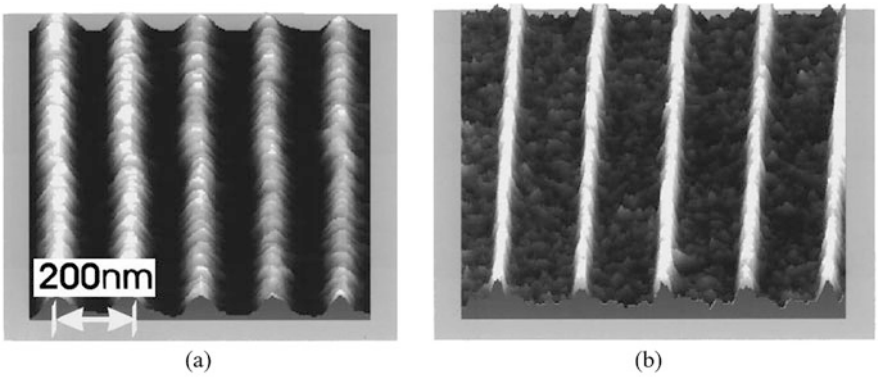


Fig. 5.11 AFM images of oxide lines produced at (**a**) 61% of humidity and (**b**) 14% humidity. (Reprint from [40] with permission)

contact mode (CM-AFM), and AFM in tapping mode (TM-AFM) for local oxidation lithography.

There are a number of factors that can influence the oxidation process, including applied bias voltage, dwelling time of tip at one location of oxidation, tip-substrate distance, and ambient humidity. Among them, the humidity plays an important role in the formation of oxide structure during tip-induced oxidation process. Figure 5.11 compares oxide line features made by AFM-induced oxidation of silicon at different humidity levels. At 61% of humidity, the oxide lines were measured around 90 nm (Fig. 5.11a). It reduced to around 22 nm at 14% of humidity level despite all other conditions being the same (Fig. 5.11b). Humidity level not only can influence the amount of water molecules supplying to the oxidation reaction but also can influence the electrical field distribution around the tip-sample region. At high humidity level, a water film is formed, which can increase the lateral distribution of electrical field, creating an equivalent "defocus" effect. Therefore, the oxide growth area becomes wider.

The growth rate of oxide can be measured by the height of oxide grown out of sample surface. For oxidation of silicon, it was found that the height of SiO_2 linearly increases with the increases of applied bias voltage (−2 V to −20 V) to the tip and the increase of dwelling time (10 ms to 1000s) [40], and this has been confirmed by an analytical model [41]. However, the oxide growth rate is high only at the beginning of oxidation (first 50 ms), then slowing down and dropping. This is

because the increase of oxide thickness has reduced the field strength across the tip and substrate silicon. The oxide growth practically stops at field strength $<10^7$ V/cm, unless the bias voltage is further increased.

Apart from silicon, a variety of other materials can also be oxidized by a scanning tip, including compound III–V semiconductors, silicon carbide, and a number of metals such as titanium, tantalum, aluminum, molybdenum, nickel, and niobium. For metallic surface, the oxidation follows a slightly different reaction [42], as expressed in Eq. (5.3).

$$M + nH_2O \rightarrow MO_n + 2nH^+ + 2ne^- \tag{5.3}$$

Unlike exposure of resists, patterned oxide can be directly used as a mask to etch into silicon, because of the high selectivity between silicon dioxide and silicon in reactive ion etching. Silicon nanowire transistors with a channel width of 4 nm were made using an AFM tip to oxidize silicon as a mask and subsequently etched into silicon to create the nanowire [43]. Local oxidation can also serve as dielectric barriers in variety of nanoscale devices such as quantum devices, nanowire field-effect transistors, graphene devices, MoS_2 transistors, photonic devices, and silicon microlens [2]. There are also other field-induced phenomena between an electrically biased tip and a sample surface, such as electrochemical etching assisted by acidic solution [44], or field-induced decomposition [45]. Though some nanoscale structures can be generated, they are generally of poor quality.

5.4.3 *Friction-Induced Modification*

Perhaps the most straightforward surface modification by a scanning probe is direct mechanical action on a sample surface. By applying a force on a tip, the friction-induced mechanical plowing on hard or soft surfaces can make grooves with widths

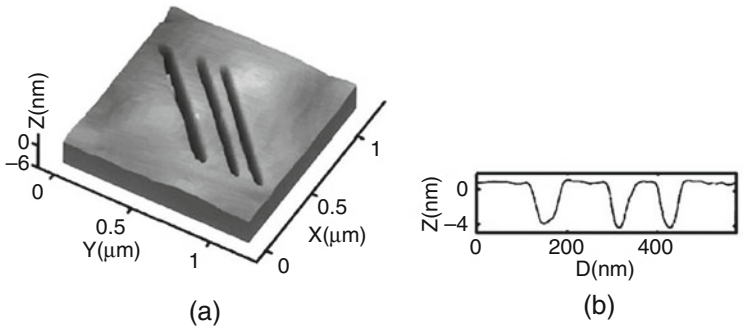


Fig. 5.12 (**a**) AFM image of scratched grooves, (**b**) depth profiles of these grooves. (Reprint from [46] with permission)

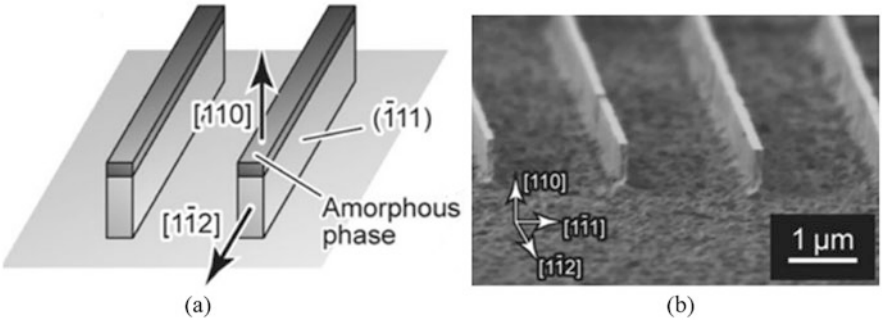


Fig. 5.13 (**a**) Friction-induced amorphous silicon as the etching mask; (**b**) KOH etching of single-crystal silicon in (112) direction

of tens of nm and depths of a few nm. Figure 5.12 shows three 500-nm-long grooves scratched by a Si_3N_4 cantilever tip on a two-layer GaSb/InAs film. Figure 5.12a is the AFM image of scratched grooves and Fig. 5.12b is the depth profiles of these grooves. The stiffness of the cantilever is 0.37 N/m and the applied forces vary from 100 to 150 nN with a scanning rate of 0.02 μm/s. The grooves produced on the GaSb (5 nm)/InAs (20 nm) substrate are about 65 nm in width and 4 nm in depth.

Friction-induced patterning by scanning probes has long been explored [47]. The resolution of mechanical scratching depends on the size of the tip and the depth of scratch. It is of the highest resolution among all the scanning probe lithography techniques described so far, because the geometry of a tip is fixed whereas electrical field or field emission electrons spread a much larger range than the boundary of a solid tip apex. Even a large tip can produce a small scratch by controlling the force and depth. For example, 3-nm-wide groove was made using a v-shaped 30-nm-radius Si_3N_4 tip with applied force of 100nN [48]. The downside of such high resolution scratch patterns is the very shallow depth. The mechanical scratch patterns made with standard AFM tips are no more than a few nanometers deep. For example, the aforementioned 3-nm-wide groove was only about 1 nm deep. Such shallow patterns are of very limited use. One way to extend the use of these mechanically scratched shallow patterns is to employ the top surface imaging (TSI) concept which has been widely used in conventional optical or e-beam lithography. One of the reported examples used a bilayer resist system consisting of a 3-nm-thick Ti film on top of a 65-nm-thick PMMA layer. AFM with Si tip scratched the 3 nm Ti film with a tip force of a few mN, and the scratched pattern is then transferred to the underneath PMMA layer by an isotropic O_2 reactive ion etching [49]. For some materials, nanoscale scratch can change the surface crystalline morphology, which leads to the change of chemical etching resistance. For example, it was found that friction forms an amorphous phase on a single-crystal silicon surface that has an etch resistance against potassium hydroxide (KOH) (Fig. 5.13a), and high aspect ratio structures were made on a (110)-oriented silicon by the so-called tribo-

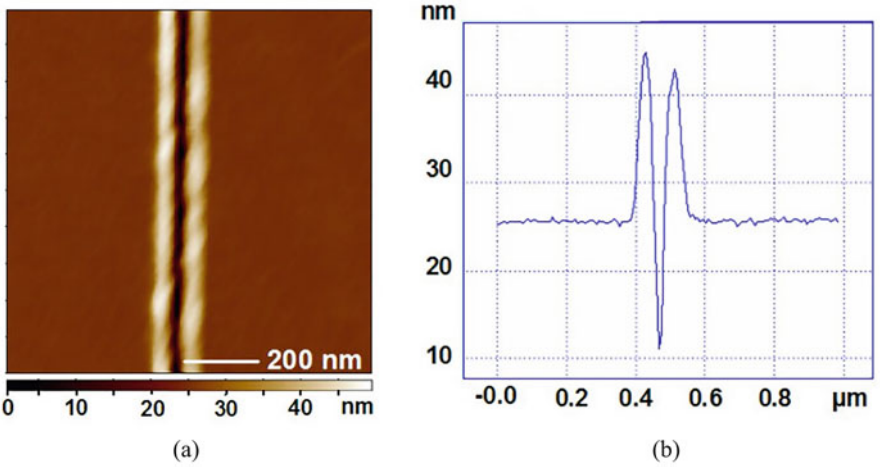


Fig. 5.14 (**a**) AFM image of plowed PMMA groove by silicon tip, (**b**) profile of the groove in (**a**). (Reprint from [52] with permission)

nanolithography (Fig. 5.13b) [50]. The friction-induced selective etching can also be applied to a variety of other materials [47].

Scratching hard surface can wear the tip and reduce its lifetime. Coating the AFM tips with hard materials or making the tip directly in hard materials such as diamond can improve its lifetime. Both carbon-coated Si_3N_4 tips and single-crystal diamond tips have been used for mechanical scratch patterning, achieving scratch depth up to 80 nm at the applied force up to 30μN [51]. The other option is to scratch soft materials such as polymers. AFM has been used for direct patterning of PMMA resist and metal nanowires can be created by subsequent lift-off process. In this case, normal silicon AFM tips can be used with applied force of about ~μN. The scratching depth of 15 nm was achieved, and more than 50 scratches could be performed with the same AFM tip because of the less damage to the tip with polymer material [52]. However, because the mechanical scratching is simply shoveling away the material from the plowing path, generation of debris along the plowing path is inevitable. For shallow scratching, debris may be less a problem. For soft materials, plowing by a scanning tip may simply squeeze the material to both sides of plowing path, leaving additional topography on the patterned surface. Figure 5.14a is the AFM image of a nanogroove formed by AFM plowing on PMMA surface. The bright edges on both sides of the groove represent higher than surface plane topography. Figure 5.14b is the profile of the groove. The bumping edges are over 20 nm higher than the PMMA surface plane, which is caused by squeezed material from the scratched trench.

If the film of polymer can move away while being plowed by an AFM tip, then the bumping edge on the sides of groove can be avoided. Such was the idea of using a strained polymer film for AFM plow patterning [53]. When an AFM tip plows through an internally strained film, the strain pulls away the film from the cutting,

leaving an open trench at the plowing path. By varying the polymer film process conditions and AFM plowing parameters, trenches with widths spanning from 40 nm to 2.3 μm have been fabricated. The final trench width is also determined by the film adhesion to the substrate, which acts against the released strain. Therefore, different trench widths can be achieved either by varying the total internal strain, for instance, by varying the film thickness, or by modifying the film adhesion to substrate surface.

Working with self-assembled monolayer (SAM) is much easier because it requires less force in the mechanical plowing process and causes less wear to the tip. A number of studies have investigated thiol-functionalized DNA on gold surface with AFM. By applying force slightly higher than what is needed for AFM imaging (30–50 nN), DNA molecules were mechanically shaved away, and the gold surface was exposed [54]. In this case the patterned SAM can act as a mask either for subsequent patterning of gold or for adding some other molecules to the patterned area. One of the experiments was to immerge the SAM-covered substrate in a solution which contained molecules different from the SAM. Where the SAM molecules were shaved away by AFM, they were replaced by the reactive molecules in the solution [55]. There is a whole range of techniques to pattern SAMs with a scanning probe [56].

5.4.4 Heat-Induced Modification

When an AFM tip is heated up, it can make dents in low melting point polymer materials. In this way, patterns of nanometer scale can be made into polymer surface. Patterning by thermomechanical indentation is different from the heated sublimation of PPA described in Sect. 5.3.3. The first thermal mechanical patterning by AFM was reported in the early 1990s [57], in which the AFM tip was heated by a pulsed laser beam to above the softening point of PMMA (120 °C). The size and depth of indentation depend on the temperature of the tip and the force of the tip pressing into the polymer. At about 10^{-7} N pressing force, the dents were typically about 100 nm across and 10–20 nm deep. The AFM tip can also be heated by building a microscale heater integrated into the cantilever beam, which makes the thermomechanical indentation system much compact. An example of such integrated cantilever has been schematically shown in Fig. 5.8. One of the prominent examples of tip-based thermal mechanical indentation was the patterned media for high density data storage. Binary data “1” and “0” are represented by dent and no dent on polymer surface. With a 40-nm-thick film of PMMA, dent sizes ranging between 10 and 50 nm have been achieved. One drawback emerged with this approach was the increased tip wear, because of the frequent contact between the Si cantilever tip and the hard Si substrate during writing. A 70 nm layer of cross-linked photoresist (SU-8) was added between the Si substrate and the PMMA film to provide a soft landing for the cantilever tip. With this configuration, dents of 40 nm in 120 nm pitch

were made in the PMMA layer, which are equivalent to 400Gb/in^2 data storage density [58].

In addition to thermomechanical indentation, highly localized heat produced by a thermal tip can change the physical or chemical properties at the tip-surface interaction point. The physical properties that can be modified by thermal tip include local amorphization of crystalline surface, local crystallization of amorphous surface, or reversion of local magnetic polarity. The chemical properties that can be modified by thermal tip include deprotection of functional surface molecular groups. For example, self-assembled monolayer which is used to protect gold surface can be deprotected by a heated tip so that the gold surface can be etched. Many chemical reactions can be activated by heat. A common strategy to obtain nanostructures of the desired material and property is to use a precursor material that is locally converted by the heated tip. One example is the reduction of graphene oxide (GO) flakes and films to obtain electrically conducting reduced graphene oxide (rGO). Thermal patterning with the heated probe locally reduces the oxygen content of the GO, which can increase the electrical conductivity by four orders of magnitude [27].

5.5 Material Deposition

5.5.1 Dip-Pen Deposition

It was long found that when AFM is used in damped air, condensed water is formed between the AFM tip and sample surface because of humidity. Depending on the humidity level and wetting property of substrate, water molecules could move between the AFM tip and substrate surface, which was difficult to control and has

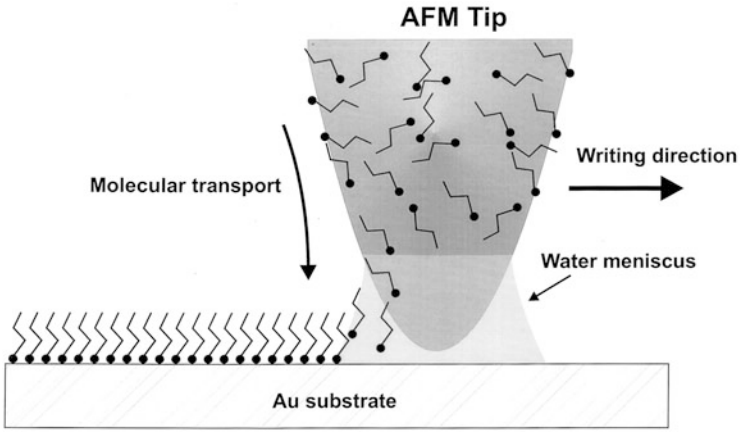


Fig. 5.15 Schematic of dip-pen nanolithography. (Reprint from [59] with permission)

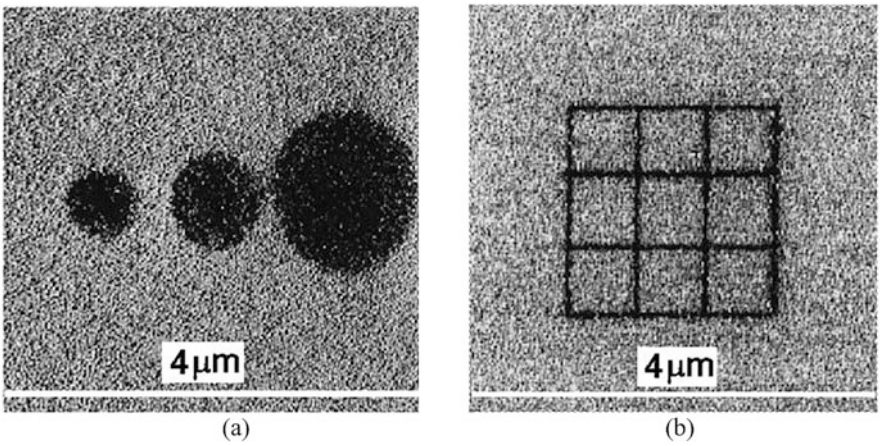


Fig. 5.16 Patterns made with DPN—(**a**) dots of different sizes representing tip-substrate contact time of 2, 4, and 16 min, (**b**) line patterns drawn by scanning tip. (Reprint from [59] with permission)

been plaguing the AFM imaging operation. In 1999, Chad A. Mirkin's group at Northwestern University of the USA utilized the phenomenon and invented the dip-pen nanolithography (DPN) [59]. Writing with a fountain pen on paper is a familiar example how ink is transported from the pen to a substrate. DPN works in the same principle. Figure 5.15 shows schematically how the ink is deposited from a probe to substrate. There is a liquid meniscus formed between the tip and substrate, which provides the route for liquid molecules to move toward the substrate surface.

It requires three conditions to complete the one-way molecule transport from tip to substrate: (1) the molecules should be able to dissolve in water; (2) there is a water meniscus existing between the tip apex and substrate; (3) the ink molecules can be self-assembled onto the substrate. In order to ensure this is one-way transport, the preferred choice of ink-substrate combination is the thiol-based molecular ink and gold substrate. Thiol group molecules are known to form stable self-assembled monolayer (SAM) on gold [60]. One of such thiol group molecules used in the first DPN experiment was 1-octadecanethiol (ODT). The factors that can influence the resolution of DPN patterning are (1) the grain size of substrate, very much like the texture of a paper when writing with a fountain pen; (2) surface chemisorption property which determines the diffusion of molecules after deposition; (3) tip dwelling time on the surface, and (4) humidity level which affects the size of meniscus formed between the tip and substrate. Figure 5.16a is the image of ODT dots deposited onto polycrystalline gold surface. The three dots represent AFM tip coated with ODT in contact with gold substrate for 2, 4, and 16 min. By moving the tip along the surface, line patterns can be drawn with ODT ink, as shown in Fig. 5.16b. Rough surface, high absorption surface, long dwelling time, and high humidity level can all contribute to large size of deposited patterns. DPN technique can routinely

produce sub-100 nm patterns, as by all other SPL techniques. However, to make features below 50 nm requires optimization of aforementioned conditions.

The three conditions to satisfy the molecule transport from tip to substrate are rather restrictive and limit the material deposition only to a small group of molecules. Many materials may not be dissolvable in water and may not be self-assembled on a specific substrate. For a general liquid ink, such as a molten polymer, there are different mechanisms for the DPN. It was found that the mass transport for a general liquid ink depends on (1) the viscosity of liquid which governs the flow rate from the tip shaft to the tip apex and (2) the interfacial energy of substrate [61]. Generally, more viscous liquid ink and lower surface energy result in smaller liquid drop size on substrate. Less viscous liquid ink on a hydrophilic surface will result in larger liquid drop size. Of course, longer dwelling time of scanning tip will create larger feature size on substrate as well.

It has been over 20 years since DPN was invented. An article was published in 2020, reviewing the evolution of DPN in commemoration of the 20th anniversary of its discovery. In the last 20 years, DPN technology has evolved into many forms such as massively parallel 2D pens, polymer pens, beam pens, hard tip with soft spring pens, etc. The pens not only can deliver inks but also can be assisted by applied electrical field to induce chemical reaction between the ink and substrate (electrochemical DPN), or the tip is heated up to melt polymers as the ink (thermal DPN), or a mechanical force is applied to the tip to scratch the surface so that ink molecules can selectively react with the material underneath (mechanical DPN), or the tips simply act as optical probes to deliver photons (optical DPN) [62].

As early as in 2003, the technology was commercialized by NanoInk, a startup company. Since then, the DPN instrument has been used in many research laboratories around the world. It was once projected as an alternative to conventional lithography because of its high resolution and direct material deposition capabilities. To tackle the low-throughput issue, massively parallel tip array tools were also developed. However, the DPN as a lithography tool has not been adopted widely by industry and it remains as a laboratory tool. The NanoInk company closed down in 2013 due to lack of profitability. Nevertheless, compared with other patterning techniques, DPN can directly add materials onto a sample surface in high resolution and as patterns. This is particularly useful for material discovery [63]. For example, multispecies of biomaterials can be added onto the same spot with high precision, allowing localized biochemical reaction. It is a high resolution version of microarray which is widely used for drug screening and biomedical testing. Other material-oriented applications such as combinatorial synthesis and heterostructure designs have been enabled by DPN [62]. A new company, TERA-print Technology, was set up in 2015, which continued the commercialization path of NanoInk with its beam pen lithography tool [64].

5.5.2 *Field-Induced Deposition*

It is well known that ions or atoms can be directly pulled out of material surface under extremely high electrical field. The phenomenon is called field evaporation [65]. The extremely high field can be obtained from a sharp tip close to a counter electrode. As mentioned earlier, STM is such a system that high field exists at the STM tip at only a few volts of bias voltage. Field evaporation of material from a STM tip was first observed in the early 1990s. In the experiment, a gold STM tip was used and negative pulse voltage of 3–4 V was applied to the tip. Gold dot was formed on substrate surface typically 15–20 nm across its base and 2–3 nm in height [66]. Material deposition was easily observed from a gold tip because gold has relatively low threshold field for field evaporation (3.5 V/A) [67], and gold surface is inert to chemical contamination. If a tungsten tip was used in combination with a gold substrate, a pit was formed instead of a mound [66]. This is because tungsten has much higher threshold in field evaporation (5.7 V/A). Instead of evaporation from tip to substrate, the substrate gold atoms evaporated onto the tungsten tip, forming a pit in the substrate.

Field evaporation alone cannot completely explain the material deposition process from a STM tip. Heating by field emission current may also be responsible for the deposition. Field emission current occurs at much lower threshold field than that of field evaporation. For gold tip, the field emission current becomes considerably large at the field as low as ~0.6 V/A [68]. High field emission current can heat up the extremity of STM tip, causing melting and flowing of tip material. This can also explain why the material deposition from the tip is sustainable despite continuous loss of material from the tip. Melting and flowing of tip material can sustain the supply of atoms toward the apex of tip so that the tip would not become blunt due to the evaporation, which is in much the same way as the working of liquid metal ion sources, as described in Chap. 4. For this reason, a negative bias to the tip is preferable because negative bias is the correct configuration for field emission to occur. Experimental observation also confirmed that negative bias of the tip produced much stable deposition [68].

Another speculation about the mechanism of tip material deposition process is the formation of nanobridge between the tip and sample surface [69]. The extremely high field at the apex of a tip can cause the tip to elongate. For metallic tip with 50 nm radius and 0.75 nm above a metallic plane, the pulling force at the tip apex is of $6x10^{-10}$ N. Coupling with the heating of field emission current, the tip can be pulled down to touch the sample surface, so that a liquid metal bridge is formed between the tip apex and sample surface. Deposition then takes place through this liquid bridge. This is very plausible because liquid metal ion sources work in the same way whereby liquid apex elongates at high field strength to form an extremely sharp emitter [70]. The existence of a liquid metal bridge was also indirectly confirmed from the measurement of resistance between STM tip and substrate. The elongation keeps the tip always in sharp geometry and maintains stable deposition over an extended period. Because of the sharpening process at high field, the

very end of the tip becomes much smaller than its original geometry. Deposited lines as small as 4–5 nm in width have been made, with minimum separation of 1 nm between the deposited features [69]. One must be aware that field-induced deposition is found much easier with metallic tip such as gold or platinum because they are soft and of low melting point and of relatively low threshold of field evaporation.

Field-induced deposition can also be done with the assistance of organometallic gases [71]. Like the local oxidation by STM in ambient humidity, organometallic gas molecules are decomposed at the high field around the extremity of STM tip, and metal atoms are deposited on substrate surface directly under the STM tip. The process is often called STM-induced chemical vapor deposition (CVD). The tip is negatively biased. There is a threshold bias voltage for different precursor gases. The threshold bias is 27 V voltage for iron carbonyl gas but 15 V for tungsten carbonyl. The deposition is normally done at emission current from 100 to 500 pA [71]. Unlike the local oxidation which has limit in the height of oxide pattern, high aspect ratio patterns can be produced by the gas-assisted deposition. The diameter of deposited dots is mostly dependent on the bias voltage. Higher voltage increases the gap between the tip and the top of the dot, which increases the effective diameter of field emission electrons, as explained earlier, thus increasing the diameter of deposited dots. The pressure of precursor gas within the range from 50 to 300 mTorr has little effect on the deposition rate or the characteristics of deposition. The deposited film contains about 50% of metal, with the rest being carbon contamination and small amount of oxygen. A range of other materials can also be deposited, based on the same principle.

AFM works the same way as STM and can transfer materials from tip to substrate at a biased voltage. As AFM tips are normally made of silicon or silicon nitride, metallic materials have to be coated onto the AFM tip in order to make the deposition happen. Once it is coated with metals, it becomes no different from a STM tip and deposition takes place under high electrical field. As AFM can work in liquid solution, it is possible to initiate electrochemical deposition using an AFM tip. Copper has been electrochemically deposited onto GaAs surface by immerging the AFM tip into a mixture of $CuSO_4$ + H_2SO_4 solution [72]. The electrodeposition of Cu is strongly dependent on the structure of the substrate, applied potential, and concentration of Cu^{2+} ions in the precursor solution. Based on the same principle, many other metallic materials can be electrochemically deposited with AFM.

5.6 High-Throughput Scanning Probe Lithography

The superior resolution capability of scanning probe lithography (SPL) has fueled widespread R&D effort in the 1990s to develop it into a competing or an alternative technology for wafer-scale nanolithography. However, it has been pointed out that four criteria dictated the viability of SPL to compete with existing lithographic technologies [73]:

- Resolution
- Alignment accuracy
- Reliability
- Throughput (wafers per hour)

The fundamental strengths of SPL are their resolution ability, as has been described in the above sections, and alignment accuracy because of their atomic-scale imaging capability. Reliability is somewhat an unproven issue as different SPL methodologies may have different reliability problems. The key nagging issue for the SPL is throughput. The throughput issue can be tackled in two aspects, increasing the scanning speed of single tip or using multiple tips.

5.6.1 High-Speed Scanning

Scanning probe microscope (SPM) is notoriously slow as an imaging tool. When SPM is compared to other real-time forms of microscopy such as the scanning electron microscope (SEM), searching for the desired feature on a sample is an extremely time-consuming process. STM or AFM normally scans at speed of 1–10 μm per second. As a nanofabrication tool, a scanning probe is even slower, because the interactions between the tip and sample material to produce nanostructures also take time.

On the instrumentation side, the increase of tip scanning speed encounters the issues of feedback control loop which has to be able to work at high frequency and of probe assembly itself which has to be able to move in high speed. The conventional piezoelectric tube which controls the height of probe has low resonance frequency and cannot respond to height variation at high-speed scanning of probe. One way to increase the speed is to operate an AFM at constant height mode so that the piezoelectric tube does not have to respond to height variation during high-speed scan. Combined with the image capturing scheme used in SEM, a scan speed of 1 mm/s was achieved [74]. Another development was to replace the piezoelectric tube with a ZnO piezoelectric actuator integrated to the cantilever. ZnO film was coated onto the cantilever beam, which bent the cantilever up to 4 μm in vertical

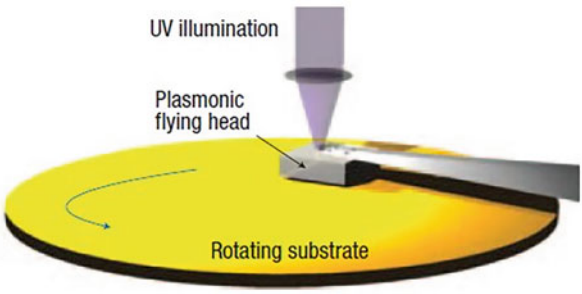


Fig. 5.17 Schematic of flying plasmonic lens for high-throughput near-field optical lithography

direction over the topographic features on a sample surface. The integrated cantilever-actuator assembly could achieve much higher resonance frequency than a piezoelectric tube. Working at constant force mode, the integrated cantilever probe achieved 3 mm/s scanning speed over a sample surface containing 2 μm vertical steps [75].

Another way to achieve high-speed scanning is to fix the probe and let the substrate move. This approach has been implemented by emulating an optical compact disk, as illustrated in Fig. 5.17 [5]. The optical read head of conventional compact disk drive is replaced by a plasmonic lens which serves as an optical probe. The substrate disk coated with a thin layer of photoresist is spinning at the speed between 4 and 12 m/s. The high-speed spinning of substrate creates an airflow which acts as an air bearing with its aerodynamic lifting force to balance the weight of plasmonic lens assembly so as to maintain 10–20 nm height above the disk surface. The plasmonic lens is made of concentric metallic rings which generate plasmonic wave when irradiated by a laser beam. Photoresist exposure by surface plasmon polariton (SPP) has been described in Sect. 2.8 of Chap. 2. The plasmonic lens can focus the SPP wave into sub-100 nm beam spot. The air bearing keeps the lens at a distance close enough to realize near-field optical exposure of photoresist. Features of 80 nm have been patterned by this method. By optimizing the plasmonic lens, 22 nm patterning is also possible [76]. A recent development increased the spin speed to 8–18 m/s at the minimum flying height of 15 nm. The circular type of plasmonic lens was also replaced by bowtie type (Fig. 5.3) to enhance the plasmonic field. With this system, feature sizes of 26–35 nm were achieved in a TeO_x film by photon-induced phase change with depth of 2–3 nm [77]. Though it looked promising, it is yet to see this type of system to be adopted by industry. Besides, the plasmonic lithography only generated the resist pattern with very shallow depth, which is really not usable directly as pattern structure, and subsequent pattern transfer by other techniques has to be carried out.

Having a high-speed scanning probe is only half of the story for high-throughput lithography. The surface modifications for the purpose of lithography take time, too, which include either exposure of resists or electrical/thermal/mechanical interaction with sample materials. Early work of tip-based lithography was using PMMA as the resist. Although it is of high resolution, PMMA is too low in sensitivity. Coupling with low efficiency of electron exposure by SPL (described in Sect. 5.3.1), much lower throughput can be expected. Use of high sensitivity resists such as chemically amplified resists is one possible route to improve the throughput. The chemically amplified resist SAL601 has been used for STM lithography, which is possible to achieve scan speed of 1 mm/s [78]. The spin-on-glass (SOG) is also highly sensitive to electron exposure. Line features with 100 nm width were achieved at scanning speed as high as 3 mm/s. The minimum 40 nm line features were obtained at scanning speed of 180 μm/s, 1 nA current, and 74 V bias voltage [79]. SOG has been commonly used in IC manufacturing as a planarization film. Electron exposure of SOG can decompose the methyl group in SOG, making the exposed region 20 times more dissolvable in buffered oxide etching (BOE) solution.

The slowest interaction between a scanning tip and a substrate material is the tip-induced hydrogen depassivation on silicon (Sect. 5.4.1), which allows the tip scanning speed of only 10–50 nm/s [36]. The fastest interaction is the heat-induced sublimation of PPA (Sect. 5.3.3). Patterning of individual pixel at the speed of 2 μs/pixel has been demonstrated in PPA by a thermal tip, which enabled high-speed linear scanning of 20 mm/s with an overall position accuracy of ±10 nm over a scan range of up to 50 μm. Pixels of 8×10^5 with feature size down to 30 nm were written in just 12 s [80].

5.6.2 *Massively Parallel Tip Array*

Even at 20 mm/s scanning speed, the mechanical movement of a probe is nowhere to match the electronic scanning speed of a focused electron or ion beam. The only way to further increase the throughput of tip-based nanofabrication is to use multiple tips and to pattern in parallel fashion. There is no fundamental hurdle and it is very much an engineering implementation issue to extend from single tip to multiple tips. The advance in silicon micromachining technology enabled massively parallel tip array to be built, including individual tip control mechanism. For example, Fig. 5.18a shows a 32 × 32 cantilever probe array chip named Millipede which was made by IBM [58]. Figure 5.18b is one of the cantilevers with a silicon tip at the top. The cantilever array was fabricated on a silicon-on-insulator (SOI) wafer, involving 11 levels of photolithography patterning and multistep reactive ion etching and metal depositions [81]. Because of the batch processing nature similar to the manufacturing of integrated circuits, the resultant 1024 cantilevers were identical and highly uniform. The Millipede was developed not for nanofabrication but for data storage. Each tip can be heated electronically and can make thermal mechanical

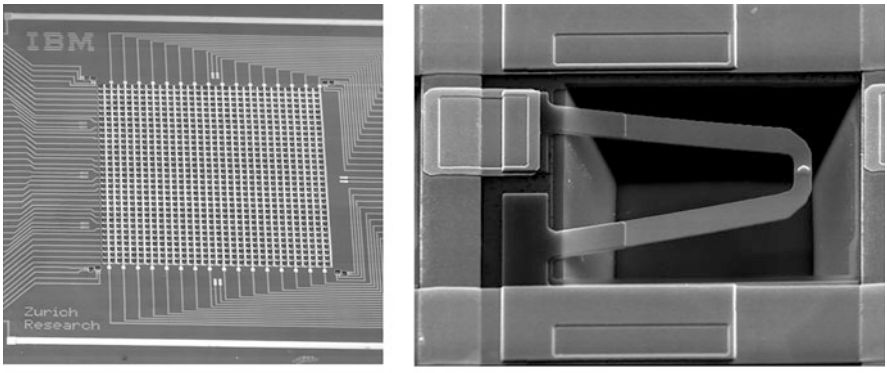


Fig. 5.18 (**a**) 32 × 32 cantilever array chip, (**b**) individual cantilever. (Reprint from [81] with permission)

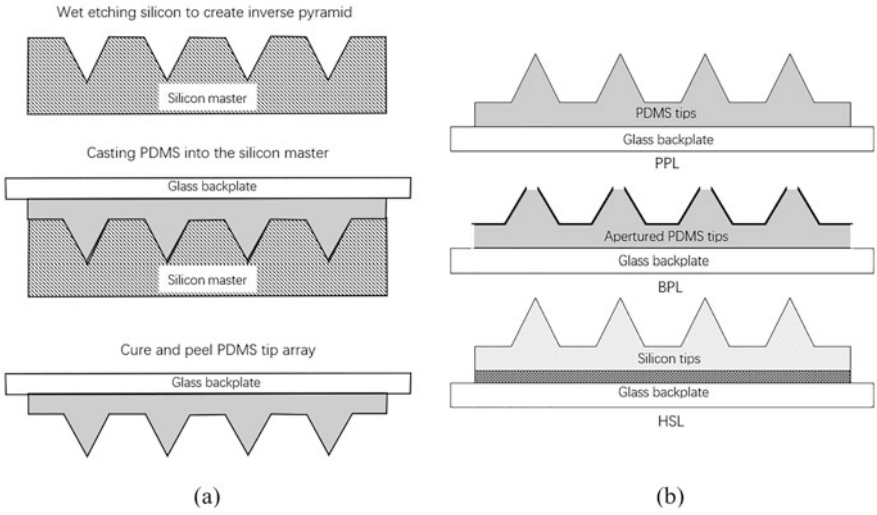


Fig. 5.19 (**a**) Fabrication of polymer pen array; (**b**) three forms of dip-pen array

indentation on a polymer surface to serve as a data storage media (Sect. 5.4.4). Although this type of data storage was not successful commercially because other technologies provided better solutions, the concept of parallel thermal tip array has been implemented for thermally converted precursor patterning [82].

The most successful high-throughput tip-based nanofabrication is realized from the evolution of DPN technology (Sect. 5.5.1). Traditional cantilever-based scanning tips cannot be scaled up to large numbers because each tip requires its own cantilever as seen in Fig. 5.18b. Chad Mirkin's group developed cantilever-free tip array based on their DPN, because the DPN is mostly for delivering liquid molecules to a sample surface and only needs a direct contact between the tip and a substrate, which can be realized by a mechanical force. There is no need to control electrical field, emission current, or heating current for individual tips. The fabrication of large array of DPN tips is relatively simple. As shown in Fig. 5.19a, pyramid-shaped pits are created in silicon by wet chemical etching; liquid elastomer (PDMS) is poured into the pits and molded into pyramid-shaped tips after curing; the slab of molded elastomer is attached to a glass plate; the vertical force applied to the soft tip array can be controlled by a piezo-scanner. Up to 28,000 polymer pens were demonstrated, and the polymer pen lithography (PPL in Fig. 5.19b) was capable to pattern features ranging in size from 90 nm to over 10 μm within 100 μm patterned area by changing the applied force and contact time [4]. In addition to the polymer pens, other forms of tip array were also developed such as the hard tip with soft spring (HSL in Fig. 5.19b) to enable pattern features smaller than 50 nm because the hard Si tips do not deform under pressure [83] and optical beam pen (BPL in Fig. 5.19b) which works as apertured NSOM (Sect. 5.3.2) to perform parallel optical lithography [84].

It has been 40 years since the tip-based nanofabrication was first discovered. Among many ways a scanning tip can provide to make nanostructures, four are better developed and more mature. They are the dip-pen nanolithography (DPN), thermal scanning probe lithography (T-SPL), hydrogen depassivation lithography (HDL), and field emission scanning probe lithography (FE-SPL). Dedicated instruments have been developed and are commercially available, such as the TERA-Fab from TERA-print using beam pens for maskless photolithography [64], NanoFrazor from Heidelberg Instruments as a table-top T-SPL system [85], ZyVector from Zyvex Labs to provide atomic-precision STM lithography based on HDL [86], and FE-SPL tool from nano analytik [87]. However, the shortcomings of tip-based nanofabrication remain, i.e., limited patterning depth, limited patterning area, and limited patterning speed. The persistent developers of these tip-based technologies hoped that one day it would play a role in nanoscale IC manufacturing. However, it may never be materialized, as can be seen in Chap. 2 on how advanced the optical lithography for IC manufacturing has progressed. Nevertheless, tip-based nanofabrication can have many applications other than competing for IC manufacturing. New patterning techniques and capabilities using scanning probes have been continuously developed, and they have played important roles in many areas of science and technology, particularly in the development of nanoscience and nanotechnology.

References

1. Ringger, M., et al.: Nanometer lithography with the scanning tunnelling microscope. Appl. Phys. Lett. **46**(9), 832 (1985)
2. Huan, H., et al.: Tip-based nanofabrication for scalable manufacturing. Micromachines. **8**(90), 1–31 (2017)
3. Kaestner, M., Rangelow, I.W.: Scanning probe lithography on calixarene towards single-digit nanometer fabrication. Int. J. Extrem. Manu. **2**, 032005 (2020)
4. Huo, F., et al.: Polymer pen lithography. Science. **321**, 1658–1660 (2008)
5. Srituravanich, W., et al.: Flying plasmonic lens in the near field for high-speed nanolithography. Nat. Nanotechnol. **3**, 733 (2008)
6. Binnig, G., Rohrer, H.: Scanning tunnelling microscopy. Helv. Phys Acta. **55**(6), 26–735 (1982)
7. Binnig, G., Quate, C.F., Gerber, C.: Atomic force microscope. Phys. Rev. Lett. **56**(9), 930 (1986)
8. Bian, K., et al.: Scanning probe microscopy. Nat. Rev. Methods Primers. **1**, 36 (2021)
9. Pohl, D.W., Denk, W., Lanz, M.: Optical stethoscopy: image recording with resolution λ/20. Appl. Phys. Lett. **44**(7), 651–653 (1984)
10. Murphy-DuBay, N., et al.: Nanopatterning using NSOM probes integrated with high transmission nanoscale bowtie aperture. Opt. Express. **16**(4), 2584–2589 (2008)
11. Folwer, R.H., Nordheim, L.W.: Proc. R. Soc. London. **A119**, 173 (1928)
12. Cui, Z., Tong, L.: Optimum geometry and space-charge effects in vacuum microelectronic devices. IEEE Trans. Electron Devices. **40**(2), 448 (1993)
13. Soh, H.T., Guarini, K.W., Quate, C.F.: Resist exposure using field-emitted electrons. In: Scanning Probe Lithography. Kluwer Academic (2001)

14. McCord, M.A., Pease, R.F.W.: Lift-off metallization using poly(methyl methacrylate) exposed with a scanning tunnelling microscope. J. Vac. Sci. Technol. **B6**(1), 293 (1988)
15. Mayer, T.M., Adams, D.P., Marder, B.M.: Field emission characteristics of the scanning tunnelling microscope for nanolithography. J. Vac. Sci. Technol. **B14**(4), 2438 (1996)
16. Wilder, K., et al.: Electron beam and scanning probe lithography: a comparison. J. Vac. Sci. Technol. **B16**(5), 3864 (1998)
17. Lenk, C., et al.: Nanofabrication by field-emission scanning probe lithography and cryogenic plasma etching. Microelectron. Eng. **192**, 77–82 (2018)
18. Krivoshapkina, Y., et al.: Low-energy electron exposure of ultrathin polymer films with scanning probe lithography. Microelectron. Eng. **177**, 78–86 (2017)
19. Kaestner, M., et al.: Nanolithography by scanning probes on calixarene molecular glass resist using mix-and-match lithography. J. Micro/Nanolithogr. MEMS MOEMS. **12**(3), 031111 (2013)
20. Betzig, E., et al.: Near-field scanning optical microscopy (NSOM) – development and biophysical applications. Biophys. J. **49**(1), 269–279 (1996)
21. Froehlich, F.F., Milster, T.D., Uber, R.: High-resolution optical lithography with a near-field scanning subwavelength aperture. Proc. SPIE. **1751**, 312–320 (1993)
22. Jin, E.X., et al.: Focused ion beam machined cantilever aperture probes for near-field optical imaging. J. Microsc. **229**, 503–511 (2008)
23. Krausch, G., et al.: Near field microscopy and lithography with uncoated fiber tips: a comparison. Opt. Commun. **119**, 283–288 (1995)
24. Ermushev, A.V., et al.: Surface enhancement of local optical fields and the lightning-rod effect. Quantum Electron. **23**(5), 435 (1993)
25. Casas, I.F., Kautek, W.: Chapter 3: apertureless scanning near-field optical lithography. In: Hu, A. (ed.) Laser Micro-Nano-Manufacturing and 3D Microprinting. Springer, Cham (2020)
26. Royer, P., et al.: Near-field optical patterning and structuring based on local-field enhancement at the extremity of a metal tip. Phil. Trans. R. Soc. Lond. **A362**, 821–842 (2004)
27. Howell, S.T., et al.: Thermal scanning probe lithography – a review. Microsyst. Nanoeng. **6**, 21 (2020)
28. Coulembier, O.: Probe-based nanolithography: self-amplified depolymerization media for dry Lithography. Macromolecules. **43**, 572–574 (2010)
29. Knoll, A.W., et al.: Probe-based 3-D nanolithography using self-amplified Depolymerization polymers. Adv. Mater. **22**, 3361–3365 (2010)
30. Lisunova, Y., et al.: High-aspect ratio nanopatterning via combined thermal scanning probe lithography and dry etching. Microelectron. Eng. **180**, 20–24 (2017)
31. Wolf, H., et al.: Sub-20 nm silicon patterning and metal lift-off using thermal scanning probe lithography. J. Vac. Sci. Technol. B. **33**(2), 02B102–02B101 (2015)
32. Eigler, D.M.S., Schweizer, E.K.: Positioning single atoms with a scanning tunnelling microscope. Nature. **344**, 524–526 (1990)
33. Ballard, J.B., et al.: Pattern transfer of hydrogen depassivation lithography patterns into silicon with atomically traceable placement and size control. J. Vac. Sci. Technol. B. **32**(4), 041804–041801 (2014)
34. Fan, P., et al.: Scanning probe lithography: state-of-the-art and future perspectives. Micromachines. **13**, 228 (2022)
35. Shen, T.-C., et al.: Atomic-scale desorption through electronic and vibrational excitation mechanisms. Science. **268**(5217), 1590–1592 (1995)
36. Ballard, J.B., et al.: Multimode hydrogen depassivation lithography: a method for optimizing atomically precise write times. J. Vac. Sci. Technol. B. **31**(6), 06FC01-1 (2013)
37. Garcia, R., Martinez, R.V., Martinez, J.: Nano-chemistry and scanning probe nanolithographies. Chem. Soc. Rev. **35**, 29–38 (2006)
38. Dagata, J.A., et al.: Modification of hydrogen-passivated silicon by a scanning tunneling microscope operating in air. Appl. Phys. Lett. **56**, 2001 (1990)

39. Fontaine, P.A., Dubois, E., Stievenard, D.: Characterization of scanning tunneling microscopy and atomic force microscopy-based techniques for nanolithography on hydrogen-passivated silicon. J. Appl. Phys. **84**(4), 1776 (1998)
40. Avouris, P., Hertel, T., Martel, R.: Atomic force microscope tip-induced local oxidation of silicon: kinetics, mechanism, and nanofabrication. Appl. Phys. Lett. **71**(2), 285 (1997)
41. Stievenard, D., Fontaine, P.A., Dubois, E.: Nanooxidation using a scanning probe microscope: an analytical model based on field induced oxidation. Appl. Phys. Lett. **70**(24), 3272 (1997)
42. Day, H.C., Allee, D.R.: Selective area oxidation of silicon with a scanning force microscope. Appl. Phys. Lett. **62**(21), 2691 (1993)
43. Martinez, J., Martínez, R.V., Garcia, R.: Silicon nanowire transistors with a channel width of 4 nm fabricated by atomic force microscope nanolithography. Nano Lett. **8**(11), 3636–3639 (2008)
44. Nagahara, L.A., Thundat, T., Lindsay, S.M.: Nanolithography on semiconductor surfaces under an etching solution. Appl. Phys. Lett. **57**(3), 270 (1990)
45. Shedd, G.M., Russell, P.E.: The scanning tunneling microscope as a tool for nanofabrication. Nanotechnology. **1**, 67–80 (1990)
46. Magno, R., Bennett, B.R.: Nanostructure patterns written in III–V semiconductors by an atomic force microscope. Appl. Phys. Lett. **70**, 1855 (1997)
47. Bingjun, Y., Qian, L.: Friction-induced nanofabrication: a review. Chin. J. Mech. Eng. **34**(32), 1–26 (2021)
48. Muller, M., et al.: Controlled structuring of mica surfaces with the tip of an atomic force microscope by mechanically induced local etching. Surf. Interface Anal. **36**, 189 (2004)
49. Hu, S., et al.: Fabrication of silicon and metal nanowires and dots using mechanical atomic force lithography. J. Vac. Sci. Technol. **B16**, 2822 (1998)
50. Noritaka, K., Noboru, M.: High-aspect-ratio structure fabrication on (110)-oriented silicon surfaces using Tribo-nanolithography. J. Nanosci. Nanotechnol. **10**(4), 2394–2400 (2010)
51. Filho, H.D.F., et al.: Metal layer mask patterning by force microscopy lithography. Mater. Sci. Eng. **B112**, 194 (2004)
52. Chen, Y., Hsu, J., Lin, H.: Fabrication of metal nanowires by atomic force microscopy nanoscratching and lift-off process. Nanotechnology. **16**, 1112–1115 (2005)
53. Jones, A.G., et al.: Highly tunable, high-throughput nanolithography based on strained regioregular conducting polymer films. Appl. Phys. Lett. **89**, 013119 (2006)
54. Zhou, D., et al.: Use of atomic force microscopy for making addresses in DNA coatings. Langmuir. **18**, 8278 (2002)
55. Xu, S., Liu, G.: Nanometer-scale fabrication by simultaneous nanoshaving and molecular self-assembly. Langmuir. **13**, 127–129 (1997)
56. Kramer, S., Fuierer, R.R., Gorman, C.B.: Scanning probe lithography using self-assembled monolayers. Chem. Rev. **103**(11), 4367–4418 (2003)
57. Mamin, H.J., Rugar, D.: Thermomechanical writing with an atomic force microscope tip. Appl. Phys. Lett. **61**, 1003 (1992)
58. Vettiger, P., et al.: The millipede-more than one thousand tips for future AFM data storage. IBM J. Res. Dev. **44**, 323 (2000)
59. Piner, R.D., et al.: Dip-pen nanolithography. Science. **283**, 661–663 (1999)
60. Xia, Y., Whitesides, G.M.: Soft lithography. Angew. Chem. Int. Ed. **37**, 550 (1998)
61. Liu, G., et al.: The role of viscosity on polymer ink transport in dip-pen nanolithography. Chem. Sci. **4**, 2093–2099 (2013)
62. Liu, G., et al.: Evolution of dip-pen nanolithography (DPN): from molecular patterning to materials discovery. Chem. Rev. **120**, 6009–6047 (2020)
63. Li, Z., et al.: Dip-pen nanolithography enabled functional nanomaterials and their applications. Adv. Mater. Technol. **6**, 2000897 (2021)
64. TERA-Print Technology. https://www.teraprint.us/
65. Muller, E.W., Tsong, T.T.: Field Ion Microscopy. Principle and Applications. Elsevier (1969)

66. Mamin, H.J., et al.: Gold deposition from a scanning tunneling microscope tip. J. Vac. Sci. Technol. **B9**(2), 1398 (1991)
67. Tsong, T.T.: Field ion image formation. Surf. Sci. **70**, 211–233 (1978)
68. Chang, C.S., Su, W.B., Tsong, T.T.: Field evaporation between a gold tip and a gold surface in the scanning tunneling microscope configuration. Phys. Rev. Lett. **72**(4), 574 (1994)
69. Houel, A., et al.: Direct patterning of nanostructures by field-induced deposition from a scanning tunneling microscope tip. J. Vac. Sci. Technol. **B20**(6), 2337 (2002)
70. Cui, Z., Tong, L.: A new approach to simulating liquid metal ion sources. J. Vac. Sci. Technol. **B6**(6), 2104 (1988)
71. McCord, M.A., Awschalom, D.D.: Direct deposition of magnetic dots using a scanning tunneling microscope. Appl. Phys. Lett. **57**(20), 2153 (1990)
72. Koinuma, M., Uosaki, K.: AFM tip induced selective electrochemical etching and metal deposition on p-GaAs(100) surface. Surf. Sci. **357–358**, 565–570 (1996)
73. Quate, C.F.: Scanning probes as a lithography tool for nanostructures. Surf. Sci. **386**, 259–264 (1997)
74. Barrett, R.C., Quate, C.F.: High speed, large-scale imaging with the atomic force microscope. J. Vac. Sci. Technol. **B9**, 302 (1991)
75. Manalis, S.R., Minne, S.C., Quate, C.F.: Atomic force microscopy for high speed imaging using cantilevers with an integrated actuator and sensor. Appl. Phys. Lett. **68**(6), 871 (1996)
76. Pan, L., et al.: Maskless plasmonic lithography at 22 nm resolution. Sci. Rep. **1**, 175 (2011)
77. Yueqiang, H., et al.: High-speed parallel Plasmonic direct-writing nanolithography using metasurface-based Plasmonic lens. Engineering. **7**, 1623–1630 (2021)
78. Marrian, C.R.K., Dobisz, E.A., Dagata, J.A.: Electron-beam lithography with the scanning tunnelling microscope. J. Vac. Sci. Technol. **B10**(6), 2877 (1992)
79. Park, S.W., et al.: Nanometer scale lithography at high scanning speeds with the atomic force microscope using spin on glass. Appl. Phys. Lett. **67**(16), 2415 (1995)
80. Paul, P.C., et al.: Rapid turnaround scanning probe nanolithography. Nanotechnology. **22**, 275306 (2011)
81. Despont, M., et al.: VLSI-NEMS chip for parallel AFM data storage. Sensors Actuators. **80**, 100–107 (2000)
82. Carroll, K.M., et al.: Parallelization of thermochemical nanolithography. Nanoscale. **6**, 1299–1304 (2014)
83. Shim, W., et al.: Hard-tip, soft-spring lithography. Nature. **469**, 516–521 (2011)
84. Huo, F., et al.: Beam pen lithography. Nat. Nanotechnol. **5**, 637–640 (2010)
85. NanoFrazor: the thermal scanning probe lithography system. https://heidelberg-instruments.com/
86. Zyvex atomic-precision lithography tool. https://www.zyvexlabs.com/
87. Field-emission scanning probe lithography tool. https://www.nanoanalytik.net/

Chapter 6
Nanoimprint Lithography



6.1 Introduction

In addition to photons or charged beams (electrons and ions) or scanning probes, lithography can be done by replication from a mold. This was the idea proposed in 1995 when Stephen Y. Chou first reported sub-25 nm holes made in PMMA polymer with an imprinting mold, and the word "nanoimprint" was coined [1]. Nanoimprint lithography (NIL) was much less complicated than optical or electron beam lithography and has been the subject of active research and development ever since it was proposed nearly 30 years ago. Vast number of papers as well as dedicated books have been published [2–4]. A wide spectrum of applications employing NIL as a patterning technique has been reported, such as quantized magnetic recording disks [5, 6], DNA electrophoresis chip [7], photodetectors [8], polarizer [9], quantum device [10], etc. As a lithographic method, any patterning processes that involve optical or e-beam lithography can be done by NIL in principle without using those expensive tools.

NIL is a parallel process, just as the optical lithography through a mask, but without the diffraction-induced limit on imaging resolution as the optical lithography has. Resolution wise, it is on par with e-beam lithography, but without the proximity effect that e-beam lithography suffers. NIL had achieved sub-10 nm resolution patterning in its early development [11]. This was particularly attractive for IC manufacturing at the time when optical lithography struggled to achieve feature size below 100 nm. NIL was indeed positioned to replace optical lithography as one of the NGL (next generation lithography) candidates for IC manufacturing by the International Technology Roadmap for Semiconductors (ITRS) in 2003 [4]. However, optical lithography moved on with its immersion, multi-patterning, and the latest EUV technology (Chap. 2). NIL never got the chance to enter the mainstream IC manufacturing. One reason is that the state-of-the-art ICs have many layers overlaying to each other, while the process of pressing a mold into the resist in NIL can cause distortions or deformations, leading to misalignments between the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_6

different layers. The other reason is that optical lithography has established mature technology infrastructure and supply chain, while there is only one company worldwide to develop the NIL for IC manufacturing over the past decades, initially the Molecular Imprints since 1999 [4, 12] and then the Canon Nanotechnology which acquired Molecular Imprints in 2014. With its continued efforts, by 2018, the company developed a cluster NIL system (FPA-1200NZ2C) capable of patterning 300-mm-size wafer at a throughput of 90 wafers an hour and sub-20 nm resolution at 3.4 nm overlay accuracy [4, 13]. Canon Nanotechnology plans to begin mass producing 3D NAND flash memory ICs using NIL by 2025.

Another area where NIL might shine was for manufacturing magnetic recording disks. The idea was to make magnetic recording bits as individual units by NIL patterning method, called bit-patterned magnetic recording (BPMR). However, novel magnetic materials were being developed rapidly, first came the perpendicular magnetic recording (PMR) and then the heat-assisted magnetic recording (HAMR). These new materials and technologies allowed magnetic storage to exceed 1 Tb/in^2 (terabits per square inch) [14], while at such data storage density, the bit size should be at least 12.5 nm if it is to be made by patterning method [15]. NIL technique has tremendous obstacles in stamp manufacturing and defects control for such small pattern size [16]. Nevertheless, recent progress in NIL for IC manufacturing promoted the renewed interest in BPMR. Western Digital, one of the world largest magnetic recording companies, demonstrated that the UV-NIL technique could make ~5.5 nm features with ~12 nm pitch using a template created from a random nanoparticle pattern, which would be equivalent to a recording density of 4 Tb/in^2 [17].

While NIL has very limited success in getting into the IC volume manufacturing and high density magnetic recording, the area where NIL has found huge opportunities in industrial scale applications is in photonics. According to a recent article in Semiconductor Engineering [18], photonics is a burgeoning industry driven by increasing global demand for light-enabled systems which are essential for a variety of products and services, including telecommunications, data networks, bio-photonics, consumer electronics, automotive, and others. These vertical markets rely heavily on optics and photonics components, such as LEDs and laser chips, optical glass, detectors and image sensors, lenses, prisms, optical filters, gratings, optical fibers, and much more, representing a global market expected to reach nearly $2 trillion by 2025. Photonics components use fewer layers than traditional ICs, which play into NIL's strength, i.e., its ability to create single layer high-resolution periodical nanostructures with excellent reproducibility and scalability on rigid or flexible substrates. NIL offers a cost-effective and efficient way to fabricate complex nanostructures below 100 nm, which is critical for sub-wavelength photonic devices such as photonic crystals, waveguides, and grating couplers.

Over the past 30 years, many forms of NIL methods have been developed. The original thermal nanoimprinting required relatively high temperature to soften the polymer and relatively high pressure to press the mold into the polymer. Heating and cooling cycle prolongs the nanoimprinting process, which is not ideal for high throughput manufacturing. Requirement for high pressure can increase the wear of

stamp and is restricted in certain applications. Although nanoimprinting based on thermal press is still being widely used today, several other forms of NIL have moved away from high temperature and high pressure processes. Notably among the derivatives of NIL is the UV cured nanoimprinting. Liquid instead of solid polymers are used, and once a liquid polymer is molded into shapes, it is solidified by photopolymerization. Some other techniques further moved away from the original concept of nanoimprinting. Techniques such as microcontact and soft lithography can no longer be called imprinting at all because the stamps are not pressed into a polymer medium to produce a surface relief pattern. In these techniques, a pattern of new material is transferred to a substrate surface by simply bringing the stamp in contact with the surface. It is a printing rather than imprinting process. However, one thing in common for all the derivatives, including the original form of nanoimprinting itself, is the replication process and the dependence on a master. This chapter intends to cover major aspects of NIL technology, particularly on patterning sub-100 nm features, the various ways to achieve this dimension scale and limitations of each technique.

6.2 Thermal Nanoimprint

Nanoimprint relying on heating and pressing was the first proposed replication process at nanometer scale [1]. It can often be confused with hot embossing, as hot embossing refers to the same replication process that relies on heating and pressing a stamp into plastics or other ductile surfaces. The general consensus is that hot embossing refers to thermal imprint of large structures which is above a micrometer or tens of micrometers, while thermal nanoimprint refers to making small structures often less than a micrometer. Therefore, the process to make plastic microfluidic devices, which are normally of tens or hundreds of micrometers in structure dimension, is called hot embossing, not nanoimprinting. Nanoimprint has much broader meaning, because some of the nanoimprint techniques may not involve heating and pressing at all, as introduced later in this chapter.

The basic process of thermal nanoimprint lithography (T-NIL) can be schematically illustrated by Fig. 6.1. A thin layer of polymer (100–200 nm thickness) coated on a flat substrate is first heated to about 50–100 °C above its glass transition temperature (T_g) so that it becomes soften. A stamp with surface relief structures is pressed into the thin polymer layer at pressure of around 50–100 bars (725–1450 psi), depending on the viscosity of molten polymer. The imprint depth is slightly less than the polymer layer thickness so that the stamp surface does not have hard contact with the substrate to prevent any damage to the stamp. The next step is to decouple the stamp from the polymer layer (demolding), while it is still soft. The temperature and pressure at different stage of imprinting are plotted in Fig. 6.2. Normally at about 10 °C below the T_g, the imprinting pressure is released. When cooling down to about 50 °C, the stamp and the imprinted sample are separated. After the separation of stamp, an impression of the stamp pattern is left

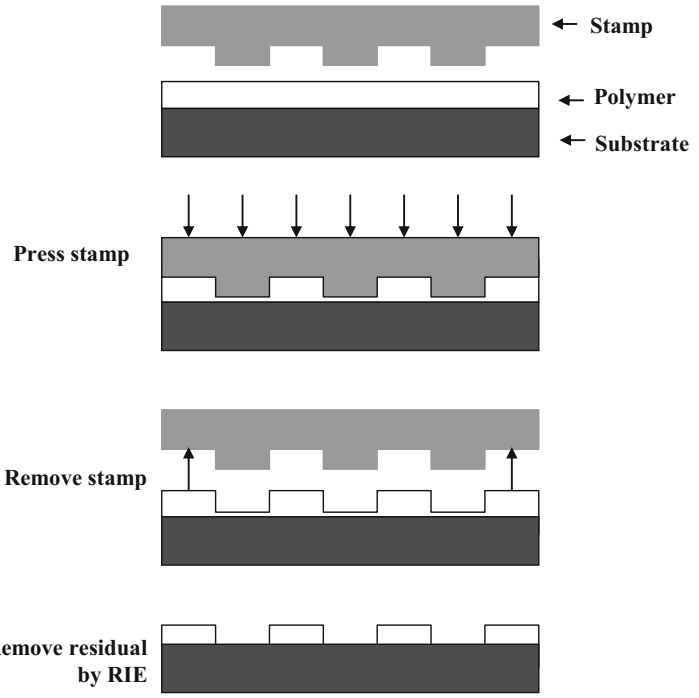


Fig. 6.1 Schematic of nanoimprinting process

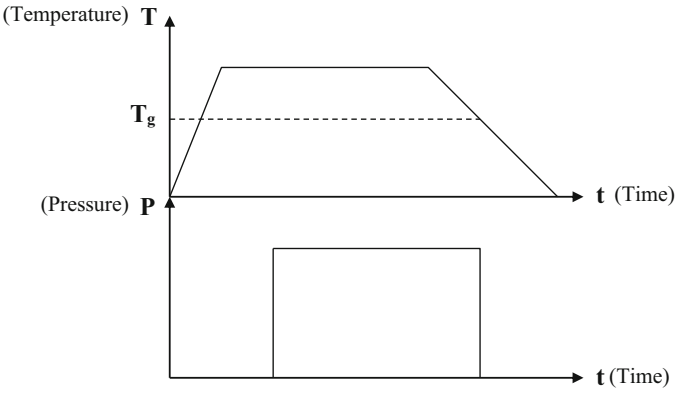


Fig. 6.2 Temperature and pressure evolution during thermal press nanoimprinting

in the polymer layer. The final step is to remove the residual polymer inside the imprinted pattern areas by reactive ion etching, to lay bare the substrate surface. The polymer patterns obtained after removal of residual polymer are in the same form as those obtained by optical lithography or e-beam lithography of resists. There are a

number of key factors which are related to faithful replication of sub-100 nm features in the thermal press nanoimprinting process, such as the stamp, the polymer, the separation of stamp and polymer (demolding), and alignment if multilevel nanoimprinting is needed. These factors are discussed in the following sections.

6.2.1 Nanoimprint Stamps

As a master in the replication process, the stamp plays the key role in replicating nanoscale structures in polymer, because the replicated structures are only as good as the master structures. For T-NIL, the stamp is made of hard material and should be able to withstand heating and pressure without significant weariness. Nanoimprint stamps can be made of many materials. However, they have to be compatible with conventional lithography process. The commonly used materials for stamps are silicon, silicon dioxide, silicon nitride, or various metals. For nanoimprinting of sub-100 nm features, the stamps are exclusively made by e-beam lithography patterning of resist and then transfer into substrate materials. Making sub-100 nm patterns by e-beam lithography has its own problems, such as proximity effect, which has been extensively discussed in Chap. 3. Once the resist patterns are made, they can be transferred by reactive ion etching (RIE) to silicon, or silicon dioxide, or silicon nitride layer on silicon. They can also be transferred by lift-off to make metal structures on silicon (see Sect. 8.3 in Chap. 8). A range of metallic materials can be used, such as gold, aluminum, chrome, or nickel. For high-resolution e-beam lithography, PMMA is the best choice of resist. However, PMMA is known to have poor selectivity in RIE. In this case, PMMA can be used as an intermediate imaging layer. After development of PMMA, a lift-off of thin chrome layer (~20 nm) is carried out to transfer PMMA pattern into chrome pattern (be aware of the change of pattern polarity). Then the chrome pattern is used as etch mask to etch into silicon, silicon dioxide, or silicon nitride up to 400 nm depth [19].

Stamps made of lift-off metal do not last very long because the metal patterns do not always stick well to silicon substrate. Stamps made of oxide or nitride are much more durable. However, silicon substrate is a brittle material. It can be easily broken under a nonuniform pressure load. The best and most durable stamp is made of nickel. The nickel stamp fabrication process is schematically illustrated in Fig. 6.3. The fabrication starts with e-beam lithography of PMMA resist on standard silicon wafer. After development of exposed PMMA, a thin metal layer (10–20 nm) of either gold or nickel is coated to cover the whole surface of PMMA layer. Then electroplating of nickel is carried out using the metal covered silicon wafer as the cathode. The electroplated nickel fills up the PMMA pattern cavities, and continuous electroplating makes the total thickness to 100–200 μm. Finally, the nickel sheet is separated from the silicon wafer by peeling off the silicon wafer if the nickel sheet is thick and strong enough or by dissolving the PMMA layer in an acetone bath. The nickel sheet can be fixed onto a metal holder plate to serve as the stamp. As nickel sheet has certain flexibility, a nickel stamp can be rolled onto a cylinder to form

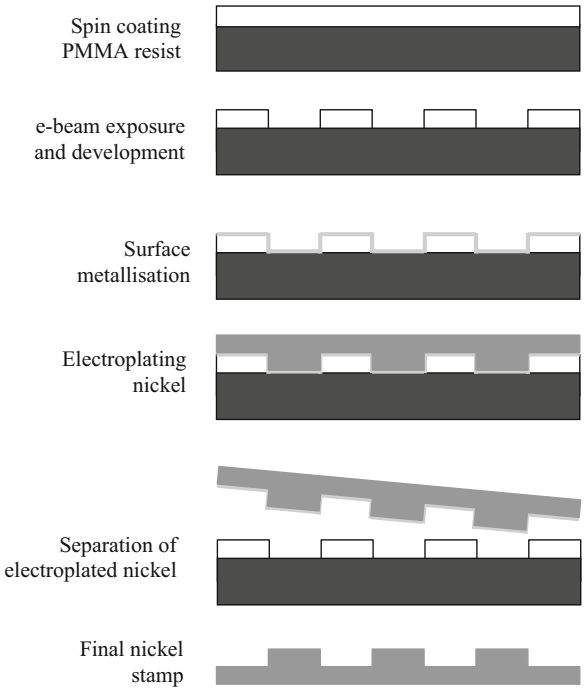


Fig. 6.3 Process for fabricating nickel stamp

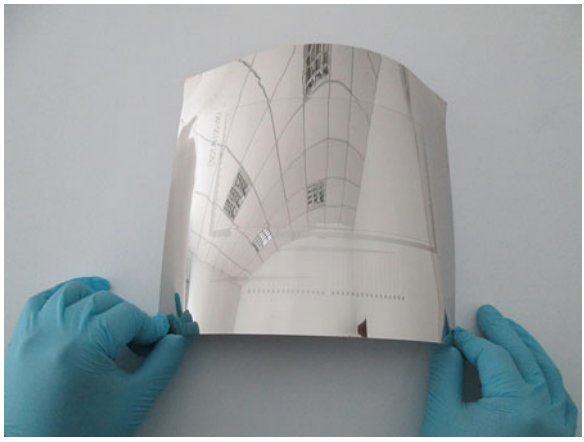


Fig. 6.4 A large size nickel stamp for roll-to-roll (R2R) nanoimprint

a roller stamp so that roll-to-roll (R2R) nanoimprinting can be carried out [20]. Figure 6.4 shows a large size nickel stamp sheet made by the author's group for R2R nanoimprinting.

Stamps can also be made without the pattern transfer step, if the patterned medium is hard enough. The HSQ resist, which has been introduced in Chap. 3, is

of such property. As previously described, HSQ resist can turn into silicon dioxide upon electron beam exposure or heating at high temperature. 20 nm line patterns with 60 nm period have been fabricated in 30-nm-thick HSQ by e-beam exposure. With a post-development baking at 500 °C, the HSQ becomes essentially silicon dioxide and is hard enough to be used as a stamp for nanoimprint [21].

It must be noted that the above stamps are only for T-NIL. Other nanoimprinting techniques may use different stamps. For example, UV cured nanoimprint has to use transparent stamps, and microcontact printing may use soft stamps. These other stamps will be described in later sections.

6.2.2 Nanoimprint Polymers

Although a high-resolution stamp is the prerequisite for high-resolution NIL, it is all down to the polymers to realize the resolution capability of the stamp. The basic requirements on a nanoimprinting polymer are low glass transition temperature (T_g) to enable low imprinting temperature, low viscosity to facilitate easy flow of polymer during filling stamp cavity process, low shrinkage after imprinting to maintain pattern fidelity, and high resistance to RIE to serve as a good mask for pattern transfer by RIE. A number of factors can influence the final outcome of nanoimprint, such as applied pressure, heating temperature, thickness of polymer layer, pattern density, and feature size on the stamp. An imprinting polymer, once heated above its T_g, has to fill the cavities of stamp by compressed flow under applied pressure. The force F required for the squeezing flow can be expressed by Eq. 6.1 [22]:

$$F = -\frac{3\pi R^4}{4h_0^3}\frac{dh}{dt}\eta_0 \tag{6.1}$$

where R is the radial flow distance, h_0 is the initial thickness of polymer layer, dh/dt is the speed of polymer layer being pressed in depth direction, and η_0 is the viscosity of polymer. The viscosity of polymer depends on both its molecular weight and heating temperature. The higher the molecular weight, the more viscous the polymer becomes. The viscosity of 350 kg/mol PMMA can be more than three orders of magnitude higher than 25 kg/mol PMMA at the same temperature [23]. With the increase of heating temperature, the viscosity of PMMA decreases exponentially. Figure 6.5 shows the viscosity of 110 kg/mol PMMA vs. temperature [24].

The above formula tells that high pressing force is required to achieve high speed of imprinting, which is also necessary when imprinting either thin polymer layer or large stamp cavity or high viscosity of polymer. High imprinting temperature is also preferred. The direct consequence of low imprinting temperature is the prolonged imprinting time at certain pressure. Imprinting, which may be completed in a few seconds at 200 °C, can take hours at 140 °C [25]. Thick polymer layer can make the

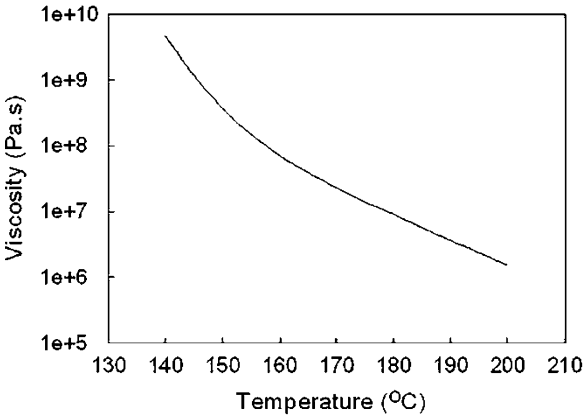


Fig. 6.5 Viscosity of 110 kg/mol PMMA vs. temperature

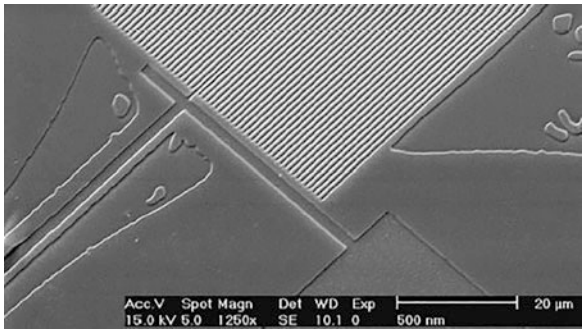


Fig. 6.6 Voids in imprinted structure due to nonuniform distribution of patterns. (Reprinted from [26] with permission)

imprinting easier. However, imprinting with thick polymer layer will leave too much residual. The final RIE step to remove the residual will take longer time. Because the RIE process is indiscriminative, it etches both the residual and the imprinted polymer pattern structure, resulting in pattern damage and distortion with excessive RIE. To have the appropriate thickness of residual layer, the amount of polymer should ideally be calculated according to the cavity volume of stamp to make sure the cavities are not underfilled in the meantime leaving the minimum thickness of residual layer.

Smaller features are easier to imprint than larger features. For example, 500 nm periodical lines can be imprinted at 100 bars pressure and 140–150 °C temperature in 5 minutes. At the same pressure and temperature, 100 µm periodical lines cannot be imprinted at all, unless the temperature rises to 200 °C [25]. In this aspect, sub-100 nm periodical features are actually easier to imprint because the squeezing flow distance of polymer is much less. The most difficult imprinting task is the imprinting of both large and small patterns. While small feature can be imprinted quickly, large feature cavities may not get filled up completely, leaving voids as shown in Fig. 6.6 [26]. The dependence of polymer filling on feature dimension and

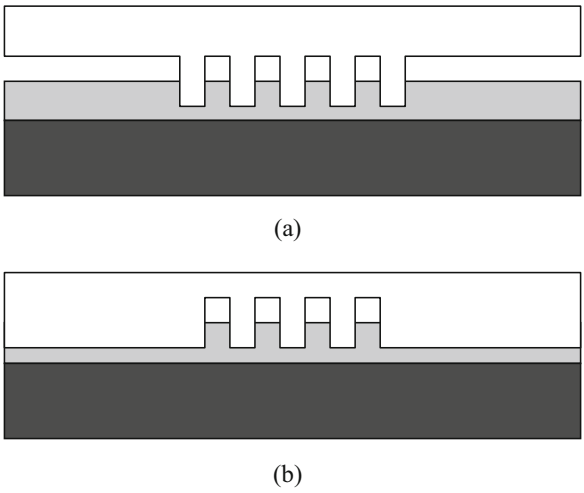


Fig. 6.7 Imprint with different stamp structures: (**a**) positive stamp with less displacement of polymer required; (**b**) negative stamp with more displacement of polymer required

density is very much like the proximity effect in e-beam lithography. In fact, it can be called the imprinting proximity effect [27].

The stamp design has serious impact on the imprinting process if the imprinting proximity effect is to be avoided. An example is shown in Fig. 6.7. For the same pattern structures, the positive stamp structure in Fig. 6.7a (protrusions) causes much less displacement of polymer during imprint, compared to the negative stamp structures in Fig. 6.7b (cavities). In general, pattern design for NIL stamps should make sure that the cavities and protrusions are uniformly distributed over the entire stamp area, to avoid large difference in polymer squeezing flow distance. This can reduce the occurrence of incomplete filling and variation of residual layer thickness. Optimizing the pressure, temperature, and imprinting time in combination with appropriate polymer material property can sometimes minimize the proximity effect and achieve uniform filling of stamp cavities with dimension from sub-100 nm to millimeter [28].

PMMA is the commonly used polymer in nanoimprint, because it is readily available in e-beam lithography. The other polymer, also widely used in NIL, is polystyrene (PS). PS is a common thermoplastic polymer and has much lower viscosity than PMMA at the same temperature [29]. The T_g of PMMA is 95–105 °C and the T_g of PS is 110 °C. These polymers are adopted for NIL because of their commercial availability, but they may not be the best suited materials. In addition to PMMA and PS, over a dozen of other polymers have also been tried for T-NIL [18, 30]. A range of dedicated polymers specifically for T-NIL are commercially available, such as the NXR-1000 series [31] and mr-I 7000R series nanoimprinting polymers [32]. These commercial T-NIL polymers are tailored to have high-resolution capability, wafer-scale uniformity, fast response and excellent flow, low processing temperature and pressure, strong adhesion to substrates, and also stability at room temperature. They are generally better in pattern transfer with RIE or liftoff.

Thermoplastic polymers can either deform plastically or elastically under heating and pressure. Plastic deformation results in permanent change of shape, which is required for imprinting. If it is elastic deformation, the polymer can recover to its original shape. There is no clear boundary between plastic and elastic deformations. It can be controlled either by the heating temperature or the pressing time. In nanoimprinting of thermoplastic polymers, the recovery can cause residual layer reshapes after the separation of stamp and sample. In the extreme case, it may recover almost to the level of the initial layer thickness. The recovery phenomenon is particularly noticeable at low imprint temperature and for large imprinted features [33].

Small imprinted features may be free from the elastic recovery problem. They however cannot avoid the internal stress induced by pressing of stamp. Free standing nanoscale polymer structures are known to be unstable. There are simulations indicating a strong reduction in the fundamental elastic moduli of the material inside a nanostructure when the width of the nanostructure decreases below 50 nm [34]. On the one hand, the polymer structures at sub-100 nm become very weak. On the other hand, large shear force can exert on the polymer structures during stamp pressing into the polymer. Internal stress can build up in these small polymer structures, weakening their rigidity and leading to structure distortion. The imprinting-induced internal stress can be manifested when the polymer structures are annealed at temperatures near the T_g of the polymer, at which temperature the internal stress can be released, if there are any, but low enough not to cause any visible flow of polymer. Pattern structure decay in the form of reduced height and enlarged base width has been observed at the annealing temperature. One example showed that line patterns of 210 nm high and 140 nm wide in 15 kg/mol PMMA decayed into 55 nm high and 180 nm wide after being annealed for 46 minutes at temperature of T_g + 15 °C [35]. The loss in height does not correspond to a proportional increase in width because the loss in height shows up as an increase of the residual layer thickness. The imprinting-induced internal stress is a general phenomenon in thermoplastic polymers. High molecular weight polymers are more viscous; therefore, they can build up more internal stress [36]. Shorter imprinting time also tends to create more internal stress [37]. With the internal stress and the elastic recovery, the pattern structure shape by T-NIL can become time dependent.

T-NIL using thermoplastic polymers needs heating and cooling which takes time. The throughput is relatively low comparing to the UV-curable resins. The advantage, however, is that these polymers are readily available at low cost and in wide variety, while UV-curable resins are generally expensive because they must contain specially designed monomer, photoacid generator, solvent to reduce viscosity, and other ingredients (see Sect. 6.4.2). The solvents normally contain volatile organic compound (VOC) which are environmentally hazardous and have to be handled with extra care [30]. Originally, the polymers for T-NIL only serve as mediums, like the resists in optical or e-beam lithography. Subsequent pattern transfer has to be performed, such as etching or deposition, in order to obtain the replicated structures in functional materials. In the past decades, many organic and inorganic functional materials have been found to be suitable for T-NIL, so that nanostructures could be

made in the materials without additional pattern transfer process [38]. Even metals can be imprinted. Direct nanoimprinting on thermally softened metals, such as Al, Ag, and Cu, has been investigated. By heating up 100-μm-thick metal foils to 500–650 °C and pressed with quartz stamp at 20 MPa pressure for 1 hour, 350 nm grating patterns were replicated in the metal foils [39].

6.2.3 Demolding

A high quality stamp and good compatible polymer imprinting medium cannot guarantee good quality nanoimprinting pattern structure unless the demolding process is also successful. Demolding is to separate the stamp from imprinted polymer structure after cooling down below the glass transition temperature of polymer. Distortion or damage of imprinted patters can happen if the polymer sticks to the stamp surface, or there is a friction drag due to roughness of stamp sidewalls, or there is a trapping of polymer due to negative slope of stamp sidewalls.

Polymer sticking to stamp is the common problem encountered in NIL because of the large surface contact area and high shear force between polymer and stamp during demolding. There are three ways to reduce the surface adhesion:

(1) To choose a stamp material which has low surface energy, such as diamond like carbon (DLC). DLC is known to have low surface energy. DLC film can be patterned to serve as the nanoimprint stamp which is not only of low polymer adhesion but also of high wear resistance [40].
(2) To mix mold release agent with polymer [41].
(3) To coat the surface of mold with an antistiction layer.

Teflon™ or Teflon-like fluorocarbon or fluorosilane release agent [26, 42] are all able to modify the mold surface to make the surface hydrophobic and low energy. Among the three methods, the antistiction coating is most widely used. The coating can be done in liquid phase or vapor phase. However, the antistiction layer become less effective as the mold cavity becomes very small, for example, a few tens of nanometers, where only a few polymer macromolecular chains are in action during imprinting.

The sidewall angle of mold has critical influence on demolding, hence on the final imprinted polymer structures. Figure 6.8 shows three different profiles of imprint stamp. It is obvious that the mold profile shown in Fig. 6.8a is preferable because it is much easier to pull out the mold from imprinted polymer layer. The demolding of the mold profile shown in Fig. 6.8b will depend on the surface roughness of sidewalls because sidewalls exert high shear force on polymer structure during demolding. The mold profile shown in Fig. 6.8c is the most undesirable profile. Imprinted polymer structures are in effect trapped inside the mold structures. Distortion and damage to the imprinted polymer structures are inevitable. Molds made from RIE or liftoff are mostly of tapered or vertical sidewall profiles and generally do

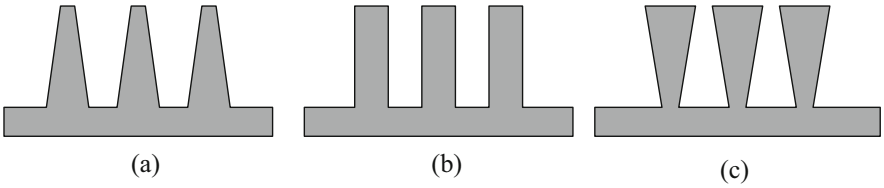


Fig. 6.8 Profiles of nanoimprint stamp: (**a**) tapered sidewall, (**b**) vertical sidewall, (**c**) undercut sidewall

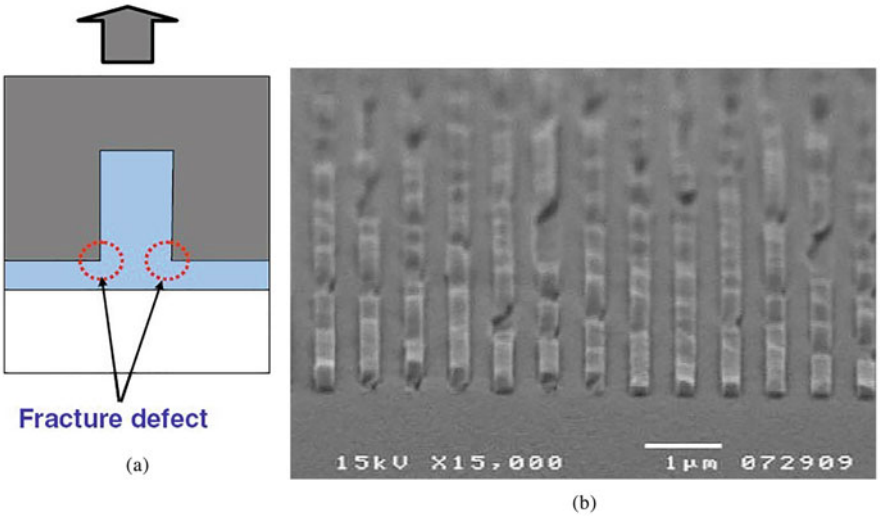


Fig. 6.9 (**a**) Illustration of stress concentration during demolding, (**b**) PMMA patterns after stamp release. (Reprinted from [44] with permission)

not have problem in demolding as long as the sidewall surface is treated with antistiction agents.

For molds with vertical sidewall profile such as the one shown in Fig. 6.8b, the higher the pattern aspect ratio (height to width), the more difficult it becomes to demold. High aspect ratio structures have large sidewall contact surface. There is a high shear stress at the vertical interface between mold and polymer. During demolding, the stress is concentrated at the base area of polymer pattern [43], as illustrated in Fig. 6.9a. As demolding takes place at temperature lower than the polymer *Tg*, polymer becomes brittle at the low temperature. Pattern fractures can happen around the base area, as the experimental result shown in Fig. 6.9b.

Using polymers with low molecular weight (M_w) to facilitate imprinting, because of its low viscosity, is generally in conflict with the requirement of safe demolding, because low M_w polymer is also low in shear modulus. This can become a serious issue when low M_w polymer is used for high aspect ratio nanoimprinting. Therefore,

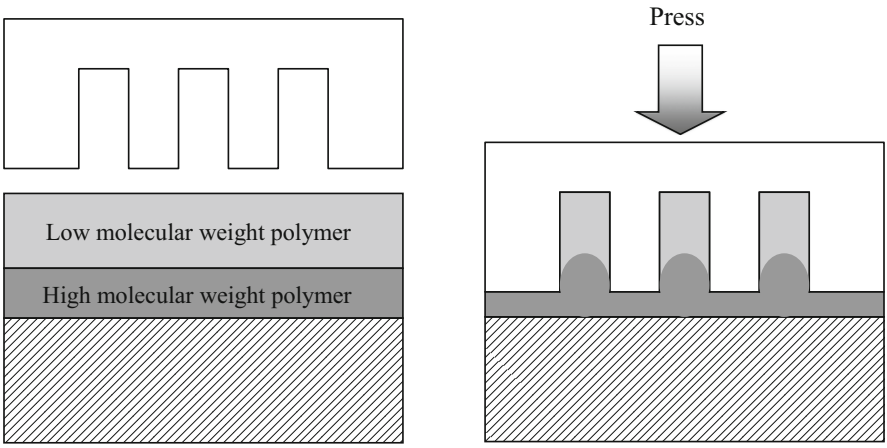


Fig. 6.10 Schematic of double-layer nanoimprinting

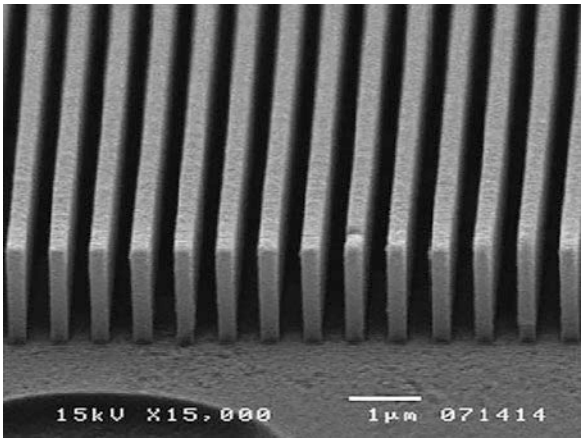


Fig. 6.11 High aspect ratio pattern imprinted using double-layer polymer process; top layer: PMMA ($M_w = 15$ k), 0.7 µm; base layer: PMMA ($M_w = 996$ k), 1.5 µm, 170 °C, 150 bars, 250 nm lines; mold pattern height: 1.3 µm. (Reprinted from [44] with permission)

for the sake of safe demolding, the pattern aspect ratio by nanoimprint rarely exceeds 3:1. For higher aspect ratio nanoimprint, one solution is to use double-layer polymers, as schematically shown in Fig. 6.10. The double-layer polymer system consists of high M_w as the base layer and low M_w as the top layer. As it is shown, when the double-layer polymers squeezed into the stamp cavity, the high M_w polymer is at the base region. Because of the high shear modulus of high M_w polymer, it can withstand the high shear force during demolding, keeping the high aspect ratio polymer pattern structure intact. Figure 6.11 shows the high aspect ratio patterns by imprinting of the double-layer polymer system.

6.2.4 Alignment

If multilevel nanoimprint patterning is required or nanoimprinting patterns need to overlay with an existing pattern on substrate surface, the overlay alignment becomes a big issue. The masters used in T-NIL are generally made of silicon or nickel. They are nontransparent. In order to align the master to an underlying pattern, an open window can be fabricated in the master, and the alignment can be carried out through the open window. However, to etch through a silicon wafer to open the alignment window, the etch-through hole has to be a few tens to a hundred micrometers. Alignment through such a large hole cannot achieve high alignment accuracy.

Another alignment technique is to fabricate alignment marks at the back of a master. Before moving the master over to a substrate, a video camera takes the image of alignment marks on the substrate which is stored in a computer. Once the master is placed on top of the substrate, an optical microscope connected to a video camera will take the picture of alignment marks on the back of the master which are compared with the substrate alignment marks already stored in computer. Both alignment mark images are displayed on a computer screen. By moving around the master relative to the substrate, while aligning the marks on computer screen, an accurate alignment can be achieved. In the early version of NIL systems when the NIL apparatus was a simple hot press without dedicated alignment facility, the aforementioned alignment was done in a mask aligner with a special holder. Once the alignment between the stamp and substrate was done, they were clamped mechanically and manually placed into the imprinter stage for imprinting. With extreme care, sub-micrometer alignment accuracy could be achieved over a 4″ substrate [45]. Later versions of nanoimprinters have the built-in online alignment facility [46]. Though the alignment accuracy is still at sub-micrometer level, the alignment process has become automatic.

The third alignment technique for NIL with nontransparent stamp is to insert a miniaturized camera between the stamp and substrate prior to pressing down the stamp [47], as illustrated in Fig. 6.12. In this way, the alignment marks do not have to be on the back of stamp, and the alignment is in situ and automatic. The alignment technique has been applied to step and stamp imprinting lithography (SSIL). SSIL

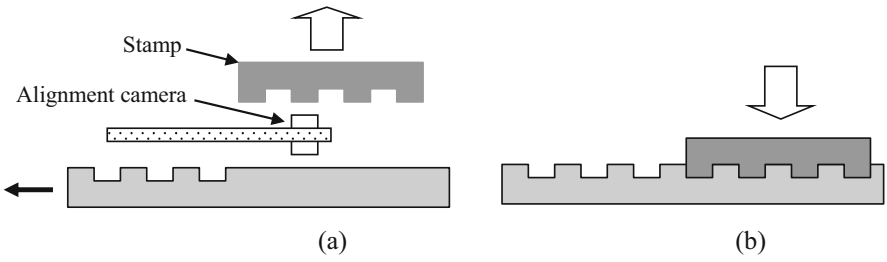


Fig. 6.12 Alignment scheme in step and stamp imprint lithography: (**a**) step, (**b**) stamp

utilizes a small stamp to imprint a large size substrate [48]. As shown in Fig. 6.12, by moving the stage which holds the substrate, the stamp imprints onto the substrate in step and repeat fashion similar to an optical stepper. Dynamic alignment is ensured in each step.

Apart from the limit in optical imaging resolution during alignment, there are other fundamental limits in the alignment accuracy for T-NIL:

1. Misalignment caused by mechanical pressure. This may be caused by nonparallelism between top and bottom platens or non-flatness of the substrate such as bowing induced by internal stress from previous processing steps. If such nonparallelism exits, a pre-aligned position before applying imprinting pressure can be offset once a pressure is applied between top and bottom platens.
2. Misalignment caused by different thermal expansions between a master and a substrate. If the master and substrate are made of different materials, their different thermal expansion coefficients can result in misalignment once they are heated to high temperature. For example, an oxide stamp to imprint on a 4″ size silicon substrate can have 100 nm/°C thermal expansion difference at the rim of 4″ silicon wafer. The effect becomes more significant when a silicon stamp is used to imprint on a GaAs substrate [49]. Therefore, for high alignment accuracy, it is imperative to avoid high temperature and high pressure.

6.3 Room Temperature Nanoimprint

Thermal press nanoimprint is not always a desirable lithography technique. In some applications, the substrate materials cannot withstand high pressure or may change property at high temperature. For example, cobalt platinum magnetic thin film is stable only below certain temperature, InP is very fragile at high uniaxial stress, and most of biological samples cannot survive high temperature and high pressure. From the throughput point of view, high temperature nanoimprinting platen needs to cool down, which consumes most of the time in an imprinting cycle. As long as the thermoplastics are used as the imprinting medium, the requirements for high temperature and high pressure are inevitable. Alterative imprinting medium is the key for the realization of room temperature nanoimprint lithography (RT-NIL), for example, using polymers with low glass transition temperature. One of such polymers is Hydrane HS2550 [50]. This polymer has glass transition temperature at 0–10 °C and melting transition point at 30 °C, which makes the polymer very suitable for low temperature NIL. There is also a report using ultrasonic energy to hammer soft metals (gold and tin) into anodic aluminum oxide (AAO) templates at room temperature [18, 51].

The breakthrough in RT-NIL was the use of hydrogen silsesquioxane (HSQ) as the imprinting medium [52]. HSQ has been introduced in e-beam lithography. It is in liquid form at room temperature and only becomes hardened when heated to dry up. Figure 6.13 shows the imprinted depth in HSQ when the resist has undergone

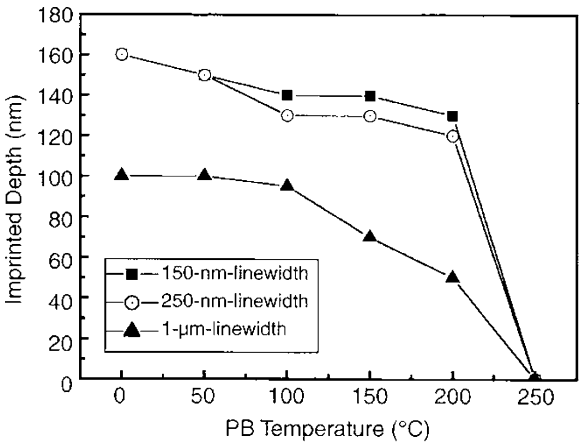


Fig. 6.13 Influence of prebaking (PB) temperature on imprinted depth in HSQ (imprinting pressure: 220 bars). (Reprinted from [53] with permission)

prebaking at different temperatures [53], indicating that it becomes harder to imprint when the HSQ layer becomes drier. Once it is dried at 250 °C, the HSQ layer cannot be imprinted at all. In practice, the HSQ layer is normally given 50–100 °C prebaking before carrying out RT-NIL, so that it is no longer "wet." The increased viscosity caused by mild prebaking can facilitate the filling of mold cavities. The plots in Fig. 6.13 also show that there is no significant increase of hardness for prebake temperature less than 100 °C. After imprinting, a post-baking is normally carried out to further dry up the HSQ layer.

HSQ is known to be difficult in spin-coating large thickness. In order to make high aspect ratio imprinted patterns or to obtain desired resist undercut profile for lift-off, multilayer RT-NIL techniques have been developed. In the multilayer resist systems, HSQ is the top layer serving as the imprinting or imaging layer. The bottom layers can be any other resists as long as they can be etched away in RIE. An example of the multilayer nanoimprinting resist system is HSQ on top of PMMA [54]. The nanoimprinting of HSQ/PMMA bilayer resist system and transfer of imprinted HSQ pattern into bottom PMMA layer by reactive ion etching (RIE) are schematically shown in Fig. 6.14. The bilayer resist system is formed by first spinning 150-nm-thick PMMA on a silicon substrate. After baking the PMMA layer in oven at 180 °C for 1 hour, the HSQ resist was spun on top of the PMMA layer at 9000 rpm spin speed, resulting in a 40 nm thickness of HSQ top layer. The HSQ/PMMA resist stack was then baked on a hotplate at 150 °C for 2 minutes and followed by oven bake of 20 minutes at 180 °C. Nanoimprinting of the bilayer resist was done by a silicon stamp at 440 bars followed by oxygen RIE to transfer the HSQ pattern into the PMMA layer. A minimum of 70 nm line features have been achieved. Normally a thin residual layer remains after imprint, which has to be removed by RIE. In the aforementioned process, such a step was not necessary. This may be due to the fact that the HSQ was so thin and had been hard baked, and there was the PMMA layer underneath to cushion the press of stamp. The stamp just broke through the thin HSQ layer into the PMMA resist. Other polymers have also been

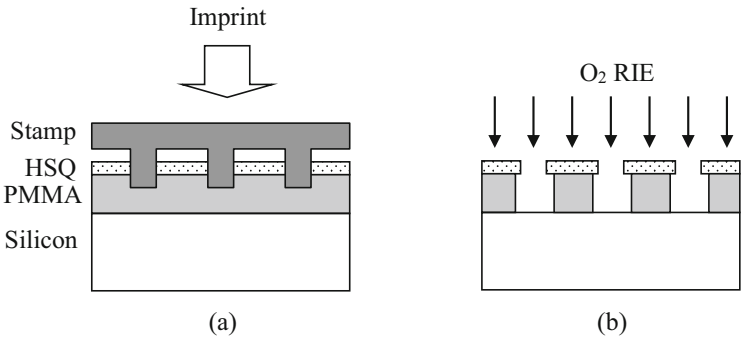


Fig. 6.14 Schematic of HSQ/PMMA bilayer resist for RT-NIL: (**a**) imprinting, (**b**) pattern transfer by RIE

used as the bottom layer, such as photoresists or poly(vinel alcohol) [53] or spin on glass (SOG) [55] or polyelectrolyte multilayer films [56]. The imprinting and pattern transfer processes are similar to what is illustrated in Fig. 6.14.

The above bilayer system offers a number of advantages over conventional T-NIL of thermoplastic polymers or RT-NIL of single layer HSQ [54]:

(1) Feature profile is independent of imprinting. High aspect ratio can be achieved in RIE of PMMA layer.
(2) The high etch resistance of HSQ to oxygen plasma allows very thin HSQ layer to be used as the imprinting layer.
(3) No need to remove residual HSQ after imprinting, resulting in better feature fidelity.
(4) The underneath PMMA can act as a buffer layer to provide a soft landing for silicon stamp, which reduces the chance of damaging the stamp.
(5) No intermediate layer is needed, because there is no intermixing between the HSQ and PMMA.
(6) There is no need to coat release agent on silicon stamp because only 40-nm-thick HSQ needs to be penetrate through. The stamp can be repeatedly used without additional cleaning process.

6.4 UV Cured Nanoimprint

Although nanoimprinting can be performed at room temperature as described in the previous section, it is still not a low pressure process. A truly room temperature and low pressure process is the UV cured nanoimprint lithography (UV-NIL). Figure 6.15 schematically illustrates the basic steps of UV-NIL. It is almost the same as the T-NIL shown in Fig. 6.1. The key differences are the use of transparent stamp and UV-curable polymer. The UV-curable polymers are in liquid form at room temperature with viscosity ranging from 50 to 200 mPa.s. Little pressure

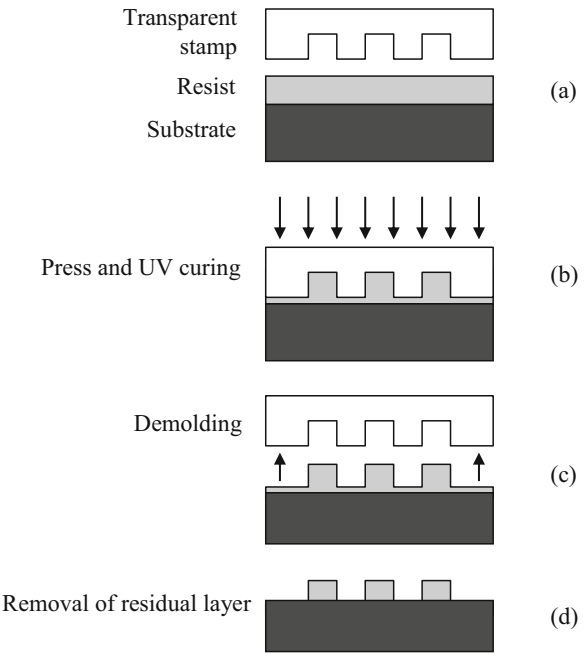


Fig. 6.15 Schematic of UV-NIL: (**a**) basic components, (**b**) UV exposure to cure the polymer, (**c**) demolding, (**d**) RIE to remove residual polymer

(< 1 bar) is needed to press the stamp into the liquid polymer. The solidification of liquid polymer is done by UV illumination through the transparent stamp. UV-NIL was first proposed in mid-1990s [57]. The technology has since gained more and more popularity as it has solved many problems associated with the T-NIL.

6.4.1 Transparent Stamps

For UV-NIL, the stamp should be transparent. In the early days of development, quartz (fused silica) substrate was the choice of stamp material. However, direct patterning resist layer on quartz by e-beam lithography met charging problem. Therefore, a thin layer (10–20 nm) of chrome is coated on the surface to act as both a discharging layer and a masking layer for quartz etching [58]. There was also a report of using indium-tin-oxide (ITO) coating as etch mask, which is both conductive and transparent [59]. Reactive ion etch (RIE) of quartz is known to be more difficult than RIE of silicon. The quartz etch rate in a conventional RIE system is only 30–50 nm/min. Even with a high power density induction coupling plasma (ICP) etcher, the etch rate is about 150–350 nm/min. Other issues such as sidewall angle and surface roughness also exist. Optimization of etching conditions is necessary for each particular stamp design in order to achieve the best quality of fabrication.

Quartz stamp is expensive in terms of both the material and the stamp fabrication process. The alternative is to fabricate stamps in polymers, as long as they are transparent to UV wavelength, and the polymer stamps can be replicated from a master silicon stamp in large quantity just like copying compact disks. As the result, the polymer stamp becomes so cheap that it can be for single use and thrown away. In this way, the master stamp can last much longer. There are two types of polymers commercially available for copying the disposable stamps. One is the Intermediate Polymer Stamp (IPS™) from Obducat AB in Sweden [18, 60], which is a polymer film that users can make their own stamps from a master stamp by a T-NIL process. The other one is the OrmoStamp® from MicroResist Technology in Germany [61], which is UV-curable organic-inorganic hybrid polymer in liquid form. The stamp made from IPS film can be used for simultaneous thermal and UV imprint, while the stamp made from OrmoStamp polymer is only for UV-NIL. A range of other polymers have also been investigated, such as a fluororesin material [62], which is >95% transparent to UV light down to 200 nm wavelength. The added benefit of this fluororesin is its natural hydrophobic surface. No antistiction coating of stamp is needed. Some fluorinated polymer stamps were repeatedly used for 200 times without obvious deterioration of surface hydrophobic property [63]. Other newly developed polymers from Optool in Sweden also do not need antistiction treatment [64]. A distinctive benefit of the polymer-based stamps is its "self-cleaning" process. If there are any particulate defects on the master stamp, the copying process can remove theses defect. As seen in Fig. 6.16, particles are simply embedded into the polymer stamp and removed during copying the polymer stamp. They would not pose any risk in the next round of copying.

As mentioned in Sect. 6.2.2, the biggest difficulty for NIL is to pattern features with large dimensional differences. While small feature cavities in a stamp can be filled quickly, large features require more volume of polymers and longer flow

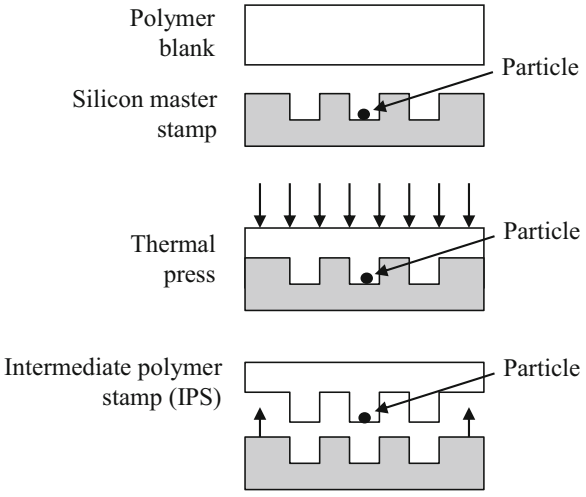


Fig. 6.16 Self-cleaning process in polymer stamp copying process

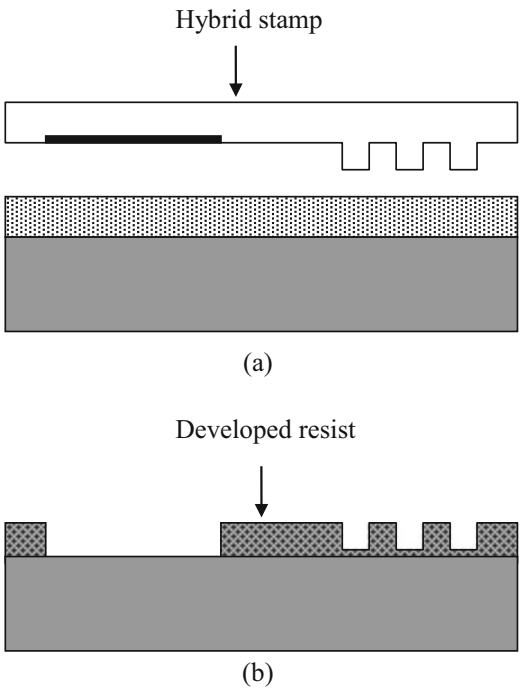


Fig. 6.17 (**a**) Hybrid stamp with opaque features, (**b**) imprinted and developed polymer structures

distance. Defects are often created due to partial filling of large stamp cavities. A hybrid stamp was developed to solve this problem, which takes advantage of UV-curable polymers, i.e., the polymer can be washed away if the UV irradiation is blocked. The hybrid stamp is schematically shown in Fig. 6.17a, which has both the transparent cavities for imprinting and opaque patterns for blocking the UV light [65]. The opaque features are formed by depositing and patterning a metal layer. When the hybrid stamp is pressed into the polymer resist and the UV light shines through the stamp, part of the UV irradiation is blocked by the opaque features on the stamp. After demolding the stamp, an additional development process is carried out so that in those areas where the UV light is blocked, the polymer resist will be washed away in developer solution, leaving both the imprinted features and the developed features as shown in Fig. 6.17b. It was demonstrated that both 350-nm-thin dense lines and 200 μm large square patterns could be made by a hybrid stamp [65].

Conventional UV-NIL always leaves an unwanted residual layer which has to be removed by RIE. A hybrid stamp shown in Fig. 6.18a can remove the residual layer without using RIE [66]. By making the end of protrusion features opaque, the UV irradiation is blocked, and no curing of polymer takes place at these areas. The subsequent removal of residual layer can be done simply by wet development in a developer solution or acetone.

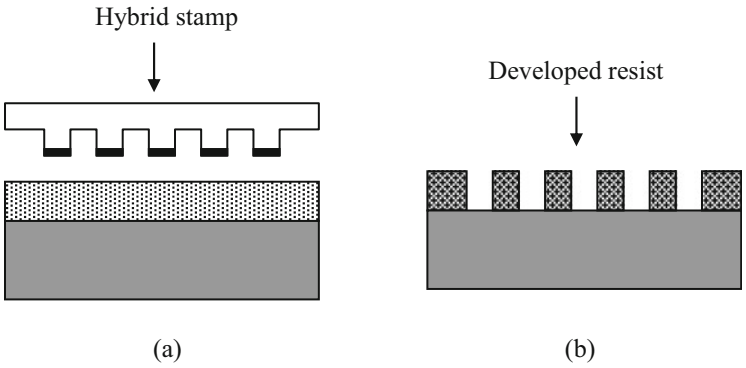


Fig. 6.18 (**a**) Hybrid stamp with opaque layer at the end of protrusion features. (**b**) Imprinted and developed polymer structures

UV-NIL does not have to use a transparent stamp. If the UV-curable polymer is coated onto a transparent substrate, a normal silicon stamp can be used, and the UV curing is done by illuminating through the transparent substrate, as schematically shown in Fig. 6.19. A thermal press NIL of SU-8 photoresist, combined with UV illumination from backside, has been able to produce high aspect ratio imprinted dense lines, as the NIL pattern shown in Fig. 6.20. Heating above the T_g of SU-8 promotes a flowable state of SU-8, which helps to fill the dense and deep cavity in silicon stamp. The extra UV curing solidifies the SU-8 and forms the final NIL pattern [67].

6.4.2 UV-Curable Polymers

The key enabler of UV-NIL is the UV-curable polymers. UV-curable polymers have existed for a long time as standard polymer products. There are UV glues and UV dental sealants. They all rely on photopolymerization process to turn from liquid or paste state into solid state upon UV light irradiation. For the UV-curable polymers to be used in UV-NIL, there are a number of specially required properties. The most important properties are as follows:

- Low viscosity so that it can fill in the stamp cavity quickly and completely via capillary force
- High resolution to replicate sub-100 nm patterns
- Fast UV curing to reduce the cycle time of UV-NIL
- Low shrinkage to ensure pattern fidelity of replication
- High dry etch selectivity to ensure reliable pattern transfer into substrate materials

These desirable properties can be obtained by appropriate polymer synthesis. A UV-curable polymer for nanoimprint commonly consists of three basic constituents:

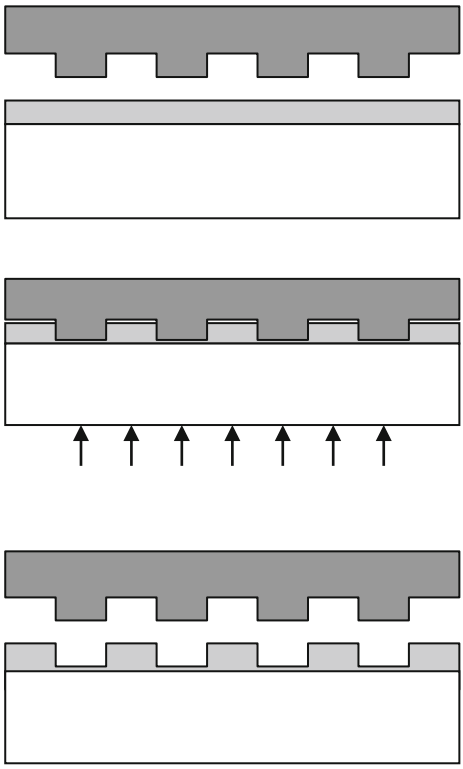


Fig. 6.19 Schematic of UV curing of polymer through a transparent substrate

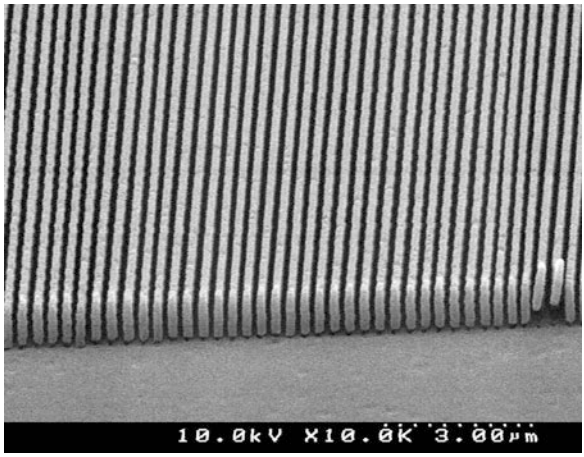


Fig. 6.20 High aspect ratio dense lines made in SU-8 photoresist by combining thermal press with UV curing nanoimprint

resin monomer, photoinitiator, and additives. To synthesize a suitable polymer is like a cooking process. Basically, different ingredients are added together so that the final mixture possesses flavors of each ingredient in various extents. To meet the above requirements, the UV-curable polymers contain vinyl ethers which are the key ingredient for photopolymerization, organic acrylates to provide the low viscosity of polymer mixture, organic cross-linker for thermal stability and mechanical strength, and photoinitiator which dissociates upon UV irradiation to form radicals which promote the polymerization process [68]. Some polymers may have additives such as silicon-containing acrylate monomers to enhance the dry etch selectivity or fluorinated compounds to facilitate mold release. For example, it is found that a fluorine based additive (1H, 1H, 2H, 2H perfluorooctyl triethoxysilane) can migrate to the surface of polymer film during spin-coating, which can lower the surface energy and create an antistiction surface to make demolding of stamp easier [69].

The viscosity of UV-curable polymers is preferably below 15 mPa.s. In practice, it is normally a few tens or a few hundreds of mPa.s. It is considerably lower than those used in T-NIL, which is at least in the order of 10^4 Pa.s [29]. Low viscosity can make the polymer filing of stamp cavity faster via capillary force. Imprinting pressure can be reduced and residual layer is thinner. The polymeric molecule size poses no limit on its high-resolution capability as the molecular chains are only 1–2 nm. The resolution of UV-NIL is primarily limited by the structure dimension of stamps. The UV curing time depends on the sensitivity and concentration of photoinitiator. Normally, the required energy for curing the polymers is in the range of a few hundreds of mJ/cm^2. The curing time is from a few seconds to a minute depending on the power of UV lamp.

One of the critical issues for UV-curable polymers is the shrinkage. UV-curable polymers, once it has been solidified, have typically 5–15% volume shrinkage [57]. The shrinkage is very much dependent on the molecular structures. It is possible to reduce the shrinkage to 3–6% by selecting high molecular weight of polymer components [70], or add a volume expanding monomer to compensate the shrinkage of conventional volume-shrinking epoxy monomers, which reduced the shrinkage to 1.86% and reduced the demolding force by 69% [71]. One thing should be noted that the polymer shrinkage is usually measured on a planar polymer film under uniform pressure and UV irradiation. The measured shrinkage in this way may not be equal to actual shrinkage in replicated pattern structures. For example, the measured shrinkage for planar film is 9%, whereas the measured dimensional variation for 100 nm line features shows no change [72]. One possible explanation is that those small features are under higher than uniform planar pressure during UV curing. The UV curing under higher pressure may result in a lower shrinkage because of the in-built internal stress.

The acrylate based UV-curable polymers have high UV sensitivity but are also sensitive to ambient oxygen. Oxygen molecules dissolved in the polymer can scavenge free radical species and thus inhibit the polymerization process at the polymer surface at the onset of exposure. This can prolong the required exposure time. For example, a UV-curable polymer has minimum UV exposure dose of 700 mJ/cm^2 in vacuum but needs 1.4 J/cm^2 when exposed in air [70]. A different

UV-curable polymer system based on cationic cross-linking of cycloaliphatic epoxies was developed to overcome this problem, because cationic polymerization is not prone to oxygen inhibition, as compared to the free radical polymerization of acrylate monomers. The polymer also exhibits a very good dry etching resistance because of the high silicon content. Furthermore, it has very low shrinkage after curing (only a fraction of that of the acrylate system) [73].

The replicated polymer structures have to be transferred to substrate. There are two stages of RIE involved. First is the etch removal of residual polymer film which is normally less than 100 nm. Second is the RIE of substrate material using the replicate polymer pattern as mask. For removal of residual film, oxygen plasma is usually used. For pattern transfer to substrate, different etching gases are used depending on the etching chemistry of substrate material. As mentioned earlier, polymers with silicon-containing components can have high resistance in RIE. Reducing the inorganic content may result in much higher etching rate during removal of residual film or pattern transfer that requires oxygen plasma. Figure 6.21 compares the pattern structures of conventional silicon-containing photoresist (SiBER from Shipley containing 19% silicon by weight) and a UV-curable polymer (containing 9% silicon) etched in oxygen plasma. The polymer with low content of silicon fared poorly in oxygen plasma etching. For this particular polymer, a completely different etching chemistry with NH_3 plasma has to be used, which exhibits selectivity of 5:1 to the etching of underneath antireflection coating.

High content of silicon may be good in pattern transfer, but making the removal residual film difficult. For example, a UV-curable polymer containing inorganic components has dry etch rate of only 10 nm/min in pure oxygen plasma. Five percent of CF_4 gas has to be added so that the plasma with O_2/CF_4 gas mixture can increase the polymer etch rate to 80 nm/min [69]. The advantage of pure organic polymer is the ease in removal of residual layer and especially in cleaning of the mold stamp if there are any residues of polymer sticking to the mold stamp. By tuning the polymer components, it is possible to formulate pure organic UV-curable polymer while

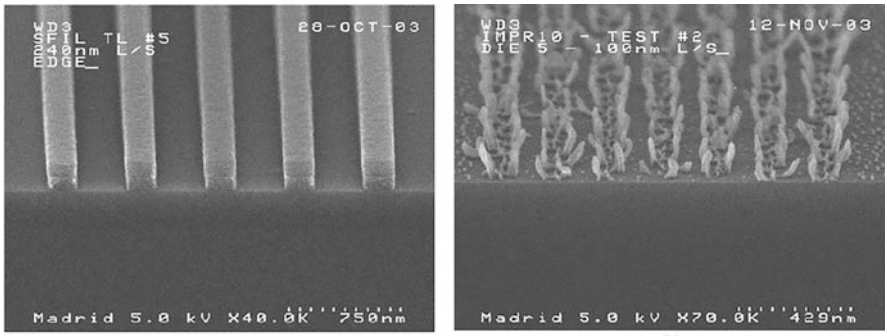


Fig. 6.21 Pattern profiles etched in oxygen plasma: (**a**) SiBER photoresist from Shipley, (**b**) a UV-curable polymer. (Reprinted from [74] with permission)

maintaining etch selectivity to silicon at around 2.5 in C_4F_8/SF_6 plasma [70]. There are commercial UV-curable polymers available such as the NXR-2000 series from Nanonex [31] and mr-NIL200 series from Micro Resist Technology [32], which have been optimized for both high-resolution UV-NIL and good etch resistance in pattern transfer.

6.4.3 *Jet and Flash Imprint Lithography (J-FIL)*

J-FIL originated from the step and flash imprint lithography (SFIL) which was first proposed in 1999 [12], and it is essentially a double-layer UV-NIL process, as schematically shown in Fig. 6.22a–d. The original UV-NIL, first proposed in 1996 [57], could only work with a thin layer, low viscosity, and UV-curable polymer film. Such a thin polymer film made it difficult for high aspect ratio pattern transfer into substrate material. In the SFIL process, a polymer resist, called transfer layer (or planarization layer), is coated onto a silicon substrate first. Then a drop of low viscosity and UV-curable silicon-containing polymer liquid, called etch barrier, is dispensed on top of the transfer layer. A transparent stamp made of quartz in the size

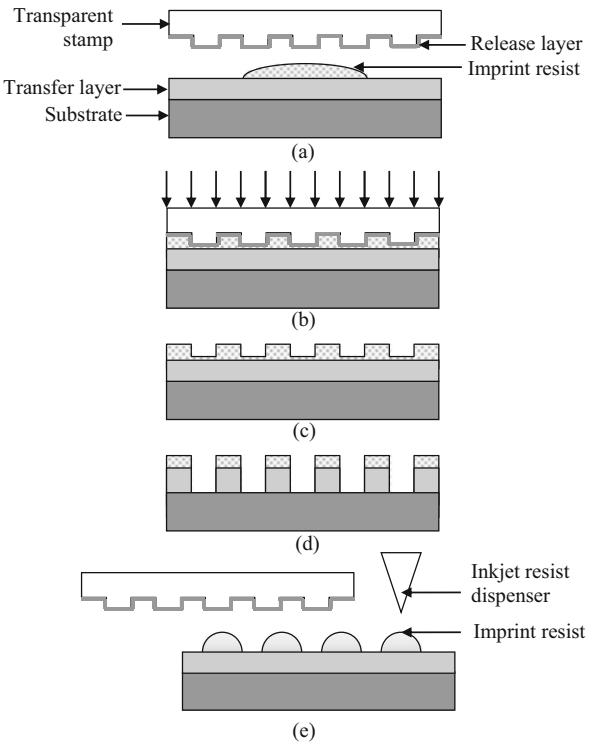


Fig. 6.22 Schematic of S-FIL: (**a**) layer structure, (**b**) imprinting and UV curing, (**c**) demolding, (**d**) RIE into transfer layer; schematic of J-FIL (**e**) intelligent resist dispensing

of 1.5 inch and coated with release layer is brought close to the transfer layer surface. The liquid polymer is then filled into the stamp cavities by capillary action. UV irradiation through the transparent stamp cures the liquid polymer and leaves an imprinted pattern of the stamp. A subsequent RIE transfers the imprinted pattern into the transfer layer which serves as the etch mask for further pattern transfer into the substrate. Same as UV-NIL, it is a room temperature and low pressure imprinting technology. Similar to the step and stamp imprint lithography (SSIL) as shown in Fig. 6.12, the SFIL imprints large size substrates by step and repeat of the stamp. Unlike the SSIL which stamps over a substrate already coated with a polymer layer, the liquid drop of etch barrier in SFIL is only dispensed at the location where the stamp is about to make an imprint.

One of the key factors preventing NIL from entering mainstream IC manufacturing is the high defects level. Apart from the defects caused by dust particles, defects can be created by insufficient filling of resist in a stamp. This is a particular issue for UV-NIL because the resist filling relies on capillary force, not on pressure and temperature as for T-NIL, and the filling characteristics are directly related to geometrical shapes and density of stamp patterns [18, 75]. On the one hand, too little resist leaves voids in the imprinted patterns. On the other hand, too much resist leaves excessive residual layer thickness. To enable an intelligent dispense of imprint resist to suit specific pattern structures, jet and flash imprint lithography (J-FIL) was developed [76]. Different from SFIL, J-FIL uses a drop-on-demand inkjet resist dispenser to deliver pico-liter-sized resist droplets in a spatially selective manner such that the dispensed resist volume is distributed to match that of the template spatial volume distribution, as illustrated in Fig. 6.22e. This ensures that all the cavities are fully filled after imprinting while leaving the residual layer extraordinarily thin (typically <15 nm) and uniform. To avoid air trapped inside the stamp cavity, the stamp is pressed in the center with a bow of $10 \sim 15$ μm using air pressure to ensure that the stamp makes first contact with the resist drops at its center and then pressed down completely, so that the resist filling starts from the center and spread out to entire stamp area [77].

Because inkjet dispense is integral to J-FIL, it is important that the resist has low viscosity between 10 and 20 mPa.s as well as high boiling point to ensure continuous jetting from inkjet printer without clogging the nozzles. Other requirements include that the resist should be easily separated from stamp without residuals adhered to the stamp and must have minimal shrinkage during curing to prevent image placement errors and to prevent significant changes to the height or linewidth of the pattern [78]. All the technical issues have been more or less solved during the 20 years of development, starting from Molecule Imprints, Inc. (MII) which was the only company determined to push nanoimprint lithography into volume IC manufacturing, until in 2014 when the company was acquired by Cannon to become Cannon Nanotechnologies. The resist structures shown in Fig. 6.23 were made by MII in early development, demonstrating the capability of this technology.

It must be admitted that J-FIL is the nearest analog to optical lithography. It is a parallel process as well as working in step and repeat mode with the stamp size of 26 mm × 33 mm same as the field size of standard 193 nm DUV lithography. The

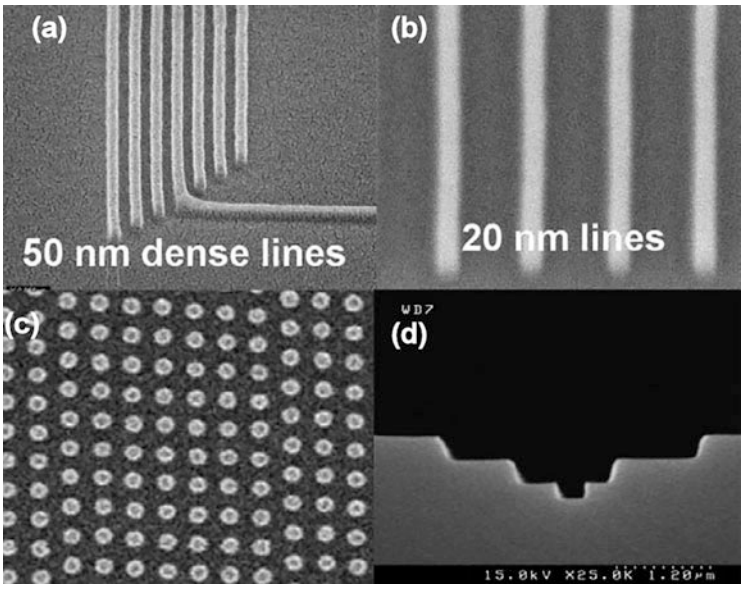


Fig. 6.23 (**a**) 50 nm dense lines, (**b**) 20 nm lines with fine line edge, (**c**) 60 nm dense dots, (**d**) imprinted 3D structure. (Reprinted from [79] with permission)

stamp is transparent, which allows high precision alignment. In fact, J-FIL has a number of advantages over optical lithography:

- Because no light wave diffraction is involved during imprinting, the only limit in resolution is the resolution of stamp structures, which at current level of e-beam lithography, as described in Chap. 3, is about 5 nm. Without light diffraction, all the optical proximity effects mentioned in Chap. 2 do not exist. Therefore, pattern in the stamp is straight design pattern without any needs of complicated OPC features.
- The line edge roughness (LER), which has become a big problem in optical lithography at sub-50 nm resolution, does not exist in J-FIL, as shown in Fig. 6.21b.
- The template can be made into 3D shape through multilevel or gray scale lithography process. Replication can be done no matter what the shape of template is. Three-dimensional pattern structure can be made easily in one step of imprint, as shown in Fig. 6.21d.
- There is no complex optical imaging system such as the 193 nm immersion optical lithography or light source such as in EUV lithography. For 20 nm dens feature lithography, the cost of process and cost of ownership for J-FIL are less than 50% of double patterning process by 193 nm immersion lithography [80].

As mentioned at the Introduction, the latest J-NIL system (FPA-1200NZ2C) from Canon Nanotechnologies performs the entire patterning in four-chamber cluster configuration, enabling a throughput of 90 wafers at 300 mm size per hour, which can almost match the throughput of mainstream 193 nm DUV lithography and has been designated for memory IC production [13].

6.4.4 Substrate Conformal Imprint Lithography (SCIL)

The stamps used for J-FIL are made of quartz which is rigid. For a large size rigid stamp, the flatness and waviness are difficult to control within a few micrometers, whereas the height of imprinted structures is not more than a few hundred nanometers. For this reason, the quartz-based stamps are never made in large size. For example, the aforementioned J-FIL stamp is 26 mm × 33 mm. With the small stamp, the imprinting of a 300 mm wafer is done by step and repeat fashion. This works for IC manufacturing because it is the size of an individual die and no field stitching is needed. For photonic applications, an optical diffractive surface can be much larger, and nanoscale structures are required to be continuous over the entire surface. By step and repeat to patch up a large area with a small stamp will inevitably encounter the field stitching issue by which a continuous feature could be disrupted due to stitching error. The large area NIL is done by a soft stamp, which is the substrate conformal imprint lithography (SCIL).

The stamp used in SCIL is made by casting liquid silicone rubber (poly-dimethyl-siloxane, PDMS) onto a glass template. The transparent PDMS stamp can be as large as 150 mm. Figure 6.24a shows such a SCIL stamp. The SCIL is done by pneumatically pushing down the soft stamp with a small pressure of 20mBar from one end and sequentially over the entire area as shown in Fig. 6.24b. In this way, no air bubbles will be trapped inside the imprinting area, and stamp cavities are filled with liquid resist by capillary force. After UV curing of resist, the stamp is peeled off sequentially by vacuum suction force. Because the stamp is soft, it can truly conform to the substrate, even over a protruded particle as shown in Fig. 6.24c [81].

Although highly conformal, a soft stamp can pose problem in patterning dense sub-100 nm features such as gratings because of pattern collapse. A composite stamp has been developed to solve the problem. The composite stamp consists of three layers: a thin layer (<100 μm) of high modulus (80 MPa) PDMS on top of 0.5-mm-thick low modulus (15 MPa) PDMS and supported by a thin glass (200 μm), as schematically shown in Fig. 6.25. The nanoscale pattern structures are made in the hard PDMS layer, which can effectively prevent pattern collapse. The soft PDMS layer acts as a cushion to provide the conformal property. The glass supporting layer is thin enough to be bendable. With the composite stamp, a grating of 25–30 nm linewidth with a pitch of 60 nm as well as 20–30-nm-diameter holes with a pitch of 50 nm can be made over 150 mm wafer size area [82].

Comparing to hard stamp UV-NIL, the soft stamp SCIL has a number of advantages [82]:

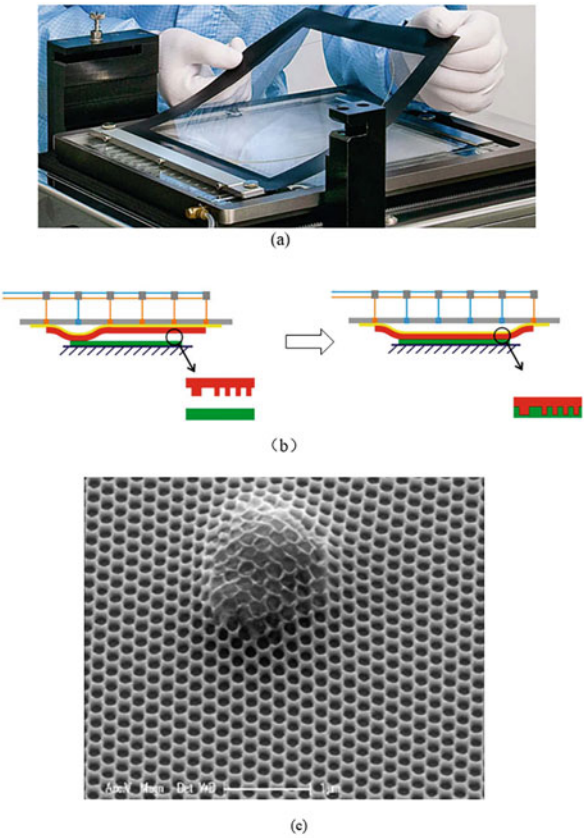


Fig. 6.24 (**a**) Soft PDMS stamp for 150-mm-size wafer SCIL. (**b**) Sequential imprinting steps. (**c**) Conformal imprinting over a particle on the substrate. (Reprinted from [81] with permission)

- Conformal contact can be made over large areas without the use of a high pressure as the flexible stamp follows the substrate curvature.
- Microscale air bubbles and partially filled features are avoided as silicone rubber has a high permeability for solvents and gasses, which allows trapped air to diffuse into the stamp.
- Particle contaminants are less problematic as the rubber can locally deform around a particle, avoiding damage to the stamp or substrate.
- Release of a rubber stamp from a rigid imprinted pattern is aided by the fact that the rubber temporarily deforms on release. This avoids damage to features in the resist or stamp and enables the replication of features with a high aspect ratio, in contrast to UV-NIL in which the aspect ratio is often limited to below 1.
- PDMS has an inherently low surface energy and is chemically nonreactive, and therefore, an anti-sticking layer is not needed.

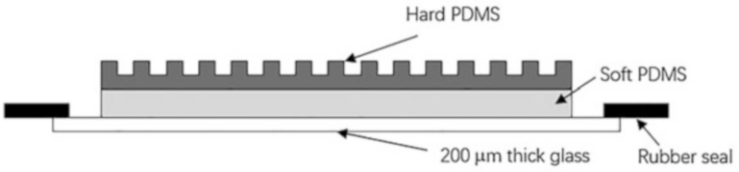


Fig. 6.25 Structure of tri-layer SCIL stamp

The SCIL was jointly developed by Philips Research and SUSS MicroTec. The technology has now become commercially available as a tool kit of SUSS mask aligner (MA/BA6 and MA/BA8). Up to 6″ full-field imprints with sub-50 nm resolution can be performed using the mask aligner with the exposure time of only 2 min if no alignment is needed (imprint 30 sec, separation 60 sec, manual wafer handling 1.5 min) [83]. As mentioned in the Introduction, the industrial scale application of NIL is for photonics. SCIL can meet the requirements of substrate size, resolution, and throughput for such applications [84].

6.4.5 Alignment Through Transparent Stamps

Apart from room temperature and low pressure processes, another key advantage of UV-NIL is its high alignment accuracy. As described in Sect. 6.2.4, thermal press imprint only has alignment accuracy of submicron because the stamp is opaque. Now UV-NIL uses transparent stamps. Alignment marks can be seen through the stamp. The alignment scheme of UV-NIL is similar to that of contact optical lithography in a mask aligner. However, it potentially has much higher alignment accuracy than either contact or projection optical lithography. Figure 6.26 schematically compares the alignment schemes for projection and contact optical lithography with that of J-FIL. In an optical projection system (Fig. 6.26a), alignment is performed in several steps: (i) to align the mask relative to the mask stage, (ii) to align the wafer relative to the wafer stage, and (iii) to align the mask and the wafer relative to each other using multiple reference mirrors. Thermal, vibrational, and mechanical noises can all contribute to alignment errors. In a contact optical lithography (Fig. 6.26b), mask and substrate have to keep a large gap in order the mask to have free movement during alignment. The large gap deteriorates the imaging of alignment mark on the substrate. UV-NIL, in particular the J-FIL, can allow much small gap ($<$0.25 μm) between the stamp and substrate because of the much small size of stamp ($\sim$ a square inch). There is a thin layer of liquid between the gap, which acts both as lubricant and damping medium. Imaging of alignment mark through such small gap is much clear and sharp using a conventional optical microscope.

As for the type of alignment marks, conventional optical mask aligners use box-in-box or cross-in-box alignment marks. Because of the similarity between contact

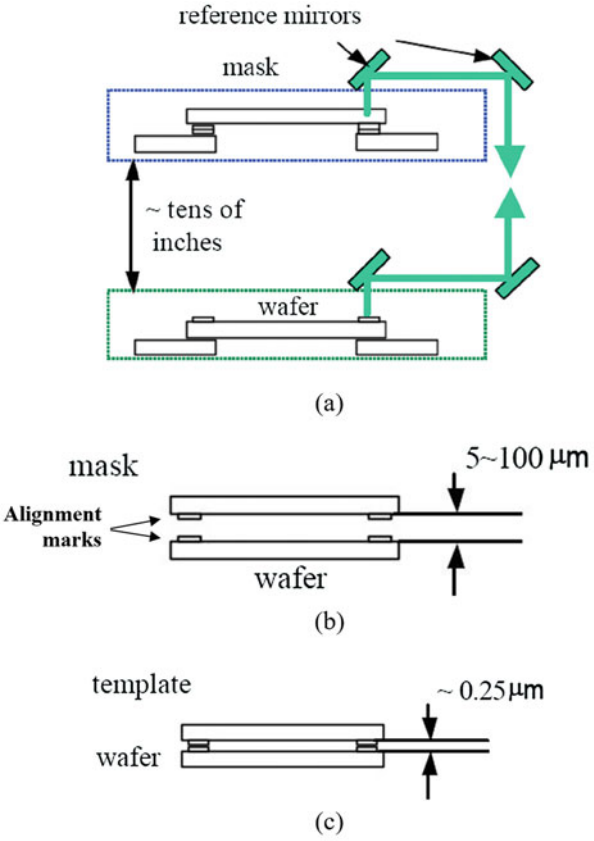


Fig. 6.26 Alignment schemes: (**a**) projection optical lithography, (**b**) contact optical lithography, (**c**) SFIL. (Reprinted from [85] with permission)

lithography using a mask and UV-NIL using a template, such alignment scheme has also been used for SCIL and is able to achieve alignment accuracy within 1 μm [82]. Further improvement in accuracy is limited by the resolution of optical microscope used for this type of alignment. A much high-resolution alignment can be achieved through interferometric alignment [86]. The interferometric alignment relies on gratings at both the stamp side and the substrate side to generate an interference pattern, called Moiré fringes. The Moiré fringes are formed when two gratings are overlaid at an angle or have slightly different periods. If the grating period on substrate is P_1 and that on stamp is P_2, then the period length of Moiré pattern can be calculated by Eq. 6.2 [87]:

$$P_M = \frac{P_2}{1 - P_2/P_1} \tag{6.2}$$

The interference pattern has much larger period length than the alignment gratings on both the stamp and the substrate. For example, if $P_2 = 1.96$ μm and $P_1 = 2$ μm, then the Moiré period length is 98 μm. Figure 6.27 is a video image

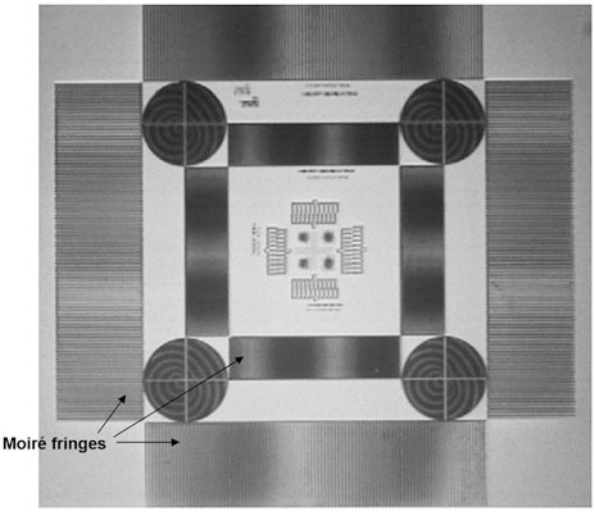


Fig. 6.27 Video image taken during alignment, showing interference dark fringes of Moiré pattern. (Reprinted from [87] with permission)

taken during alignment of stamp and substrate. The formation of Moiré pattern is seen as the dark fringes. If there is a misalignment Δ between stamp and substrate, then the Moiré period will shift a distance of $d = \Delta(P_M/P_2)$. A 150 nm misalignment between stamp and substrate will show up in Moiré pattern by 7.5 μm shift, which is much easy to see from an optical microscope. By optimizing the design of gratings, the Moiré fringe method can magnify position errors between the two gratings up to a factor of ~3000. Sub-10 nm alignment accuracy has been achieved for SCIL over 30×30 mm^2 area [82]. For J-FIL, the production tool from Cannon Nanotechnologies has achieved sub-5 nm overlay accuracy across 300 mm wafer [13]. Theoretically, the method has the ultimate alignment accuracy of ~0.5 nm [86]. Although Moiré fringes can achieve sub-100 nm alignment accuracy, it is normally used in fine-tuning step. Conventional box-in-box alignment technique is still used at coarse alignment step.

6.5 Reverse Nanoimprint

So far, the nanoimprint processes introduced are all using a structured stamp to press into a polymer-coated rigid flat substrate. To imprint on a flexible polymeric substrate, the coating cannot be done easily. For T-NIL, heating of the polymer layer can also soften the polymer substrate. In such situations, the imprinting can be done reversely [88]. The reverse nanoimprint process is schematically illustrated in Fig. 6.28. Instead of spin-coating polymer onto a substrate, the polymer is spun-coated onto a structured stamp to fill up all the stamp cavities (Fig. 6.28a). A short time soft bake is normally carried out to evaporate solvent in the polymer. The molded polymer structure is then transferred to a flat substrate either by thermal press

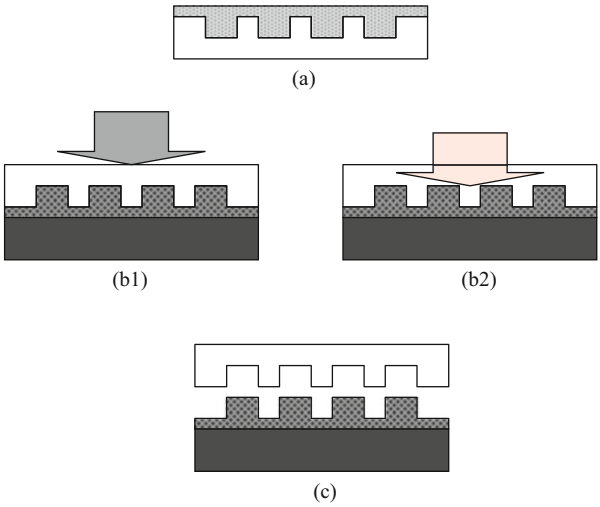


Fig. 6.28 Reverse nanoimprinting process: (**a**) spin-coating stamp with polymer to fill up the mold cavities; (**b1**) reverse imprinting onto substrate by thermal press; (**b2**) reverse imprinting onto substrate by UV curing; (**c**) demolding

(Fig. 6.28b1) if the polymer is PS or PMMA or by UV curing (Fig. 6.28b2) if the polymer is UV curable and the mold is transparent [89]. The UV curing can also be performed from substrate side if the substrate is transparent and the mold is opaque [90]. A pressure is needed to achieve intimate contact between the polymer-coated stamp and substrate surface. After demolding, the polymer structures molded by the stamp is transferred onto the substrate (Fig. 6.28c), which is equivalent to that obtained by normal imprint. The key condition to the success of reverse nanoimprint is to make sure that the surface energy of substrate is greater than that of the mold, so that the structured polymer can easily come off the mold but firmly stick to the substrate. Ferroelectric polymer nanostructures have been successfully made by reverse nanoimprint over large areas on flexible substrates [91].

Because polymer is poured into mold cavities, instead of pressing into mold cavities, high aspect ratio replication can be more easily achieved than conventional thermal press nanoimprint. Figure 6.29 shows replicated PMMA pillars with 17.5:1 aspect ratio by pouring the PMMA into a mold [92]. There are two difficulties facing the high aspect ratio replication: one is how to get the polymer into such deep and narrow cavities, and second is how to pull the high aspect ratio polymer structure out of the mold. The first difficulty can be overcome by using a polymer liquid with very low surface tension. Polymers diluted in solvent such as toluene or chloroform is suitable for this purpose. For example, PMMA powder is dissolved in toluene in 1: 40 weight ratio. A further vacuum baking can be performed to ensure there are no air bubbles trapped inside the deep cavities. The demolding is less a problem in reverse nanoimprint than in thermal press nanoimprint, because polymer molecules get into the mold cavity purely by gravity and capillary action. There is no external force

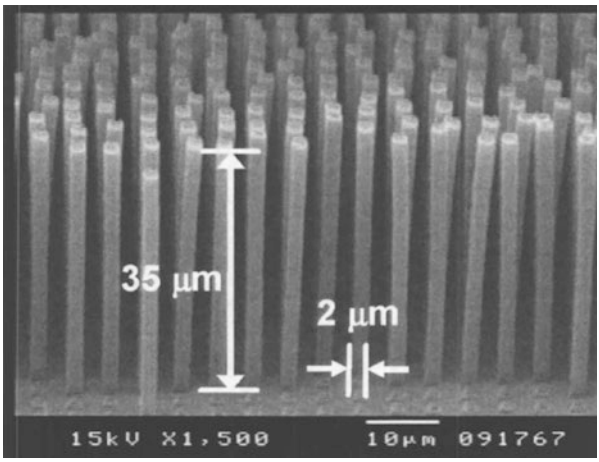


Fig. 6.29 High aspect ratio PMMA pillars replicated by cast molding. (Reprinted from [92] with permission)

applied to squeeze the polymer into cavity. Therefore, there is no shear stress created in the molded polymer structures. With proper antistiction treatment of mold surface, demolding of polymer structure should be easier, even when they are of high aspect ratio. Besides, UV curing process or drying of solvent can result in some shrinkage of polymer volume, which also facilitates the demolding process. Though the polymer structures shown in Fig. 6.29 are much bigger, the same process has been demonstrated to be able to mold sub-100 nm pattern structures in PMMA and polycarbonate (PC) [93].

The reverse imprinting can be done either on a flat substrate or a substrate that already has some structures, as schematically shown in Fig. 6.30. An example of such transfer is to use a SiO_2 mold which has 350-nm-deep grating with a 700 nm period. To assist polymer transfer to the substrate, the mold is coated with 1H,1H,2H,2H-perfluorodecyl-trichlorosilane, which results in a very low-energy surface. The mold is then spin-coated with a polymer solution whose concentration is chosen so that the mold surface is planarized, having a surface profile of oscillating waves with the maximum peak height less than 100 nm. After spin-coating, the mold is soft baked at 80 °C for 5 min to remove residual solvent. The coated mold is then pressed against a pre-patterned SiO_2 substrate in a preheated hydraulic press. The typical imprinting pressure is 5 MPa. Cooling is initiated after 5 min, and the pressure is sustained until the temperature has decreased to below the glass transition temperature (Tg) of the polymer. The mold is then separated from the patterned substrate, leaving polymer patterns on top of the protrusions on the substrate. By repeating the process, three layers of PMMA patterns were stacked up, as shown in Fig. 6.31 [94]. Reverse imprint was also demonstrated using a polymer (PVA) mold to transfer up to six layers of polymer (M-PDMS) nanostructures onto a flexible polymer (PET) substrate [95].

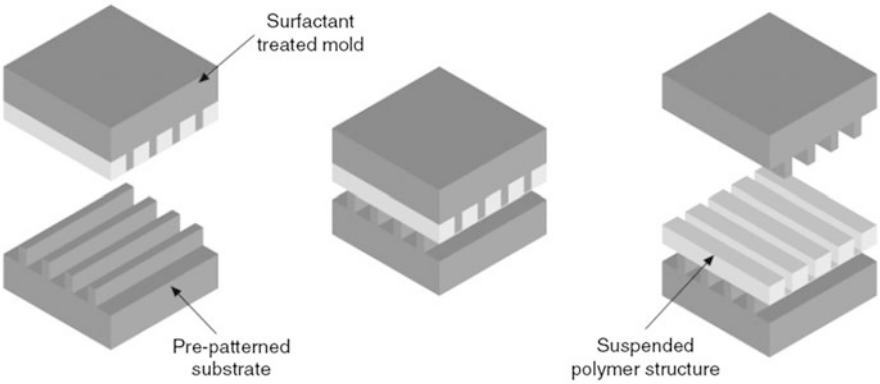


Fig. 6.30 Schematic of reverse imprint onto a pre-patterned substrate

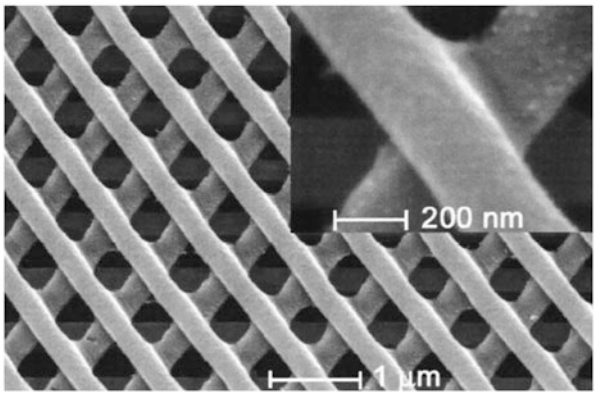


Fig. 6.31 Three layers of PMMA pattern structures are imprinted by reverse nanoimprint. (Reprinted from [94] with permission)

6.6 Soft Lithography

The previous section has shown that molded polymer pattern structures can be transferred from the stamp to a substrate by pressing the stamp in contact with the substrate. The same technique can be used to transfer liquid "ink" from the stamp to substrate. In fact, the concept of stamping ink onto a surface was proposed before the advent of nanoimprint. In 1993, Kumar and Whitesides reported using an elastomer stamp, made by casting PDMS (polydimethylsiloxane) into a master mold, to transfer alkanethiol ink to a substrate coated with gold thin film [96]. The alkanethiol ink pattern then acts as a gold etching mask to transfer the ink pattern into a gold thin film pattern on a substrate. The process is schematically illustrated in Fig. 6.32.

The PDMS (from Dow Corning with trade name Sylgard 184) is first mixed with curing agent (Sylgard curing agent 184) in 10:1 volume ratio. A vacuum de-airing is normally performed to remove air bubbles introduced during mixing. The mixed

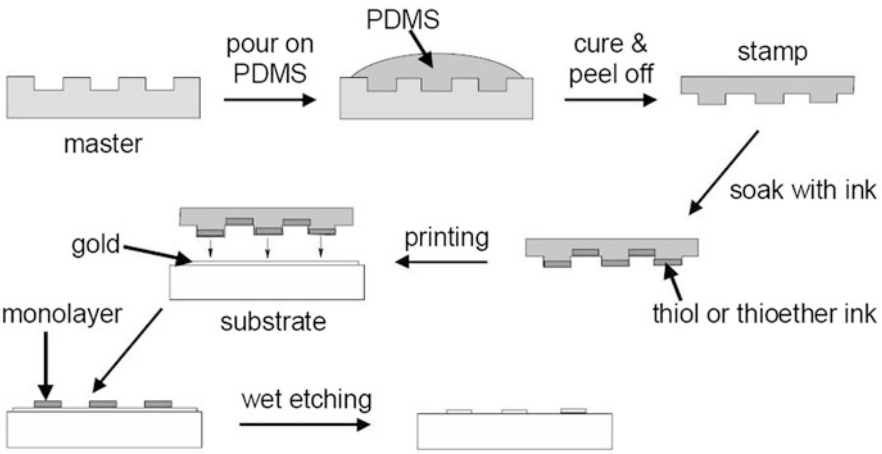


Fig. 6.32 Schematic of soft stamp fabrication and soft lithography processes

liquid is then cast onto a patterned master surface. The curing can be done at room temperature (23 °C) for 24 hours, or 65 °C for 4 hours, or 100 °C in 1 hour, or 150 °C in 15 minutes. After curing the rubber-like PDMS sheet is peeled off and soaked with thiol solution (inking). The stamp is then brought into contact with a substrate coated with thin gold film. Upon contact, the thiol ink is transferred from the stamp protrusion surface to the gold film. The thiol pattern serves as an etching mask during gold wet etch. In this way, the stamp pattern is transferred into a gold thin film pattern on the substrate. The initial work reported in 1993 might not be perfect. The choice of PDMS for making the stamp might just be because it was cheap and easy to shape into a stamp. However, the concept has been taken up by many research groups around the world in the following decades [97]. Though the word "soft lithography" was coined in 1998 [98], many variations of soft lithographic technique have been developed since, such as microcontact printing (μCP), replica molding (REM), microtransfer molding (μTM), micromolding in capillaries (MIMIC), and solvent-assisted micromolding (SAMIM) [99]. It has become an important replication technique parallel to the NIL technology, particularly for biotechnology applications [100, 101].

6.6.1 Soft Stamps

Soft stamps are the key components in the whole array of soft lithography techniques. Why soft? It is because a soft stamp allows intimate contact between stamp surface and substrate surface even when the substrate surface is rough or non-flat. Unlike nanoimprint in which a stamp embosses into a polymer layer by external pressure or capillary force as has been introduced in Sects. 6.4.3 and 6.4.4, soft

lithography transfers the ink pattern simply by surface contact. PDMS was the earliest and still is the primary choice of material. PDMS has already been introduced as a stamp material in substrate conformal imprint lithography (SCIL) (Sect. 6.4.4). Here some of the properties are reiterated as they are particularly well suited for soft lithography [102]:

1. PDMS is deformable and can easily make conformal contact with a nonplanar surface at micrometer scale. The elastic nature and low surface energy (~22 mJ/m^2) also allow it to be peeled off easily. On the other hand, the surface energy can be increased by oxygen plasma treatment so that surface binding becomes irreversible.
2. PDMS is chemically inert. It is resistant to many diluted acids, stable against many solvents, and a good etch mask for RIE of silicon. It is also optically transparent, thermally stable, and permeable to gases.
3. PDMS stamps are easy to make at low cost.

However, the very fact of softness of PDMS also causes several drawbacks in soft lithography, particularly when trying to pattern sub-100 nm features. As schematically shown in Fig. 6.33, the low tensile modulus (~1.8 MPa) can lead to pattern structure collapse if h/a is too large or to pattern sagging if w/a is too large [103]. One can imagine that for sub-100 nm stamp structures, the standing features are weaker than larger features and the depth in the stamp is shallower, all of which can lead to the aforementioned collapse and sagging. In fact, PDMS stamp structures tend to merge or collapse during inking and printing for structure dimension smaller than 1 μm [104].

The problem of soft PDMS stamp in printing sub-micrometer patterns is best demonstrated in Fig. 6.34, where the printed dense dot array from 250 nm down to 80 nm is compared using stamps with two different hardness [105]. One stamp was made of commercial PDMS (Sylgard 184) with modulus of 2 MPa, and the printed dot array is shown in Fig. 6.34a. The other stamp was made of modified PDMS material with modulus of 9.7 MPa, and the printed dot array is shown in Fig. 6.34b. The Sylgard 184 stamp can only print out 250 nm dot array, while the hard PDMS can print features down to 80 nm. Even for the largest 250 nm dots, soft stamp is

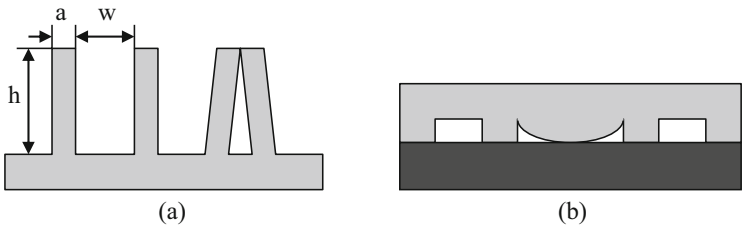


Fig. 6.33 Distortion of PDMS stamp: (**a**) pattern collapse due to high aspect ratio and close packed structures; (**b**) pattern sagging due to large span and shallow cavity features

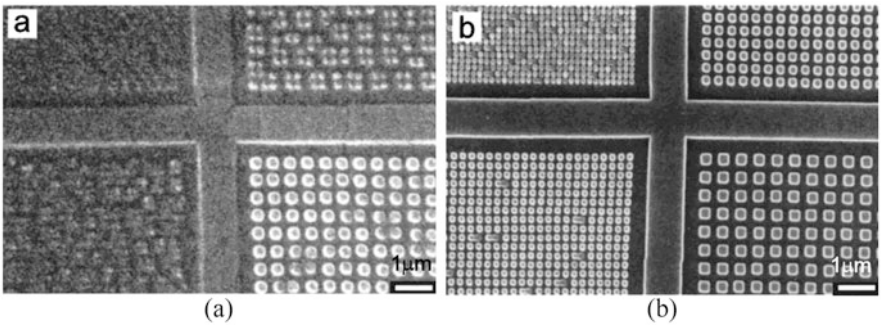


Fig. 6.34 Comparison of dot array pattern printed by (**a**) normal PDMS stamp and (**b**) hard PDMS stamp; dot sizes range from 250 nm (lower right quadrant) down to 80 nm (upper left quadrant), and structure height is 100 nm. (Reprinted from [105] with permission)

printed out in round shape, while hard stamp is printed in square. In general, commercial Sylgard 184 is only suitable for pattern structures above 500 nm [102].

In addition to the short range distortion, there is long range distortion in the form of pattern displacement error. Unsupported stamps made of Sylgard 184 showed nonuniform pattern displacements up to 20 μm between two successive prints over areas of $5 \times 5\text{cm}^2$ [104]. The use of a rigid backplane can overcome such long range distortion problem [106], such as the composite stamp shown in Fig. 6.25, which is of tri-layer structure as described in Sect. 6.4.4. The glass back plate can increase the overall rigidness to avoid long range distortion, while the patterned hard PDMS layer can eliminate the short range distortion. A minimum of 23 nm features have been imprinted with the tri-layer flexible stamp on a 4″ size substrate [107]. Other hybrid mold approach using modified polymer as hard structured layer in combination with PDMS also demonstrated 50 nm UV-NIL results [108]. It is also possible to tailor the length of molecular chains in polyurethane acrylate (PUA) so that the tensile modulus of stamp can be tuned to either 320 MPa (hard stamp) or 19.8 MPa (soft stamp). Such a stamp can be either used for hard press imprint or soft contact imprint [109].

Apart from mechanical instability problem, there are other problems associated with PDMS soft stamps: PDMS shrinks about 1% after curing; it can swell when absorbing some solvents such as toluene and hexane [99]. All the above problems make it much harder for the soft lithography to achieve resolution as good as the nanoimprint techniques described in previous sections.

6.6.2 Microcontact Printing

Microcontact printing (μCP) is an alternative name for soft lithography. μCP is possible because of two essential components: one is the PDMS stamp and another is the alkanethiol self-assembled monolayer (SAM) system. The previous section has

emphasized on the soft aspect of contact printing. This section will focus on its printing characteristics. PDMS has an important property, apart from those mentioned previously, that is, the permeability to some solvents. As a stamp, PDMS has to absorb inks first before it can stamp an ink pattern onto a substrate. The ink is normally alkanethiol dissolved in a solvent, and PDMS can absorb the solvent like a sponge. In this case, aqueous ink (water based) is not suitable because alkanethiol has low solubility in water and PDMS does not soak up water well (low permeability). The commonly used solvent is ethanol. Although ethanol is not the only solvent that can permeate PDMS, it causes less swelling in PDMS than others. The optimum concentration of thiols in ethanol is between 0.1 and 1 mM, depending on the resolution of stamp structures. Normally, low concentration is preferable for printing small features [99].

Inking of a PDMS stamp is normally done by placing a drop of thiol solution (about 1 μL per mm^2 area of stamp) onto the patterned stamp for 30 seconds, as shown in Fig. 6.35a. The drop is then blown away with N_2, and the stamp is dried for 10 seconds under a continuous stream of N_2. The inked stamp is then used to print within ~15 seconds after inking. Another "wet inking" method is to dip the PDMS stamp into thiol solution for a few minutes and blow dry under a N_2 stream, as shown in Fig. 6.35b. Although there was attempt to attach a reservoir of thiol solution to a PDMS stamp to provide constant supply of thiol ink, it did not succeed because of the swelling of PDMS stamp which was constantly soaking up with the ethanol solvent [110].

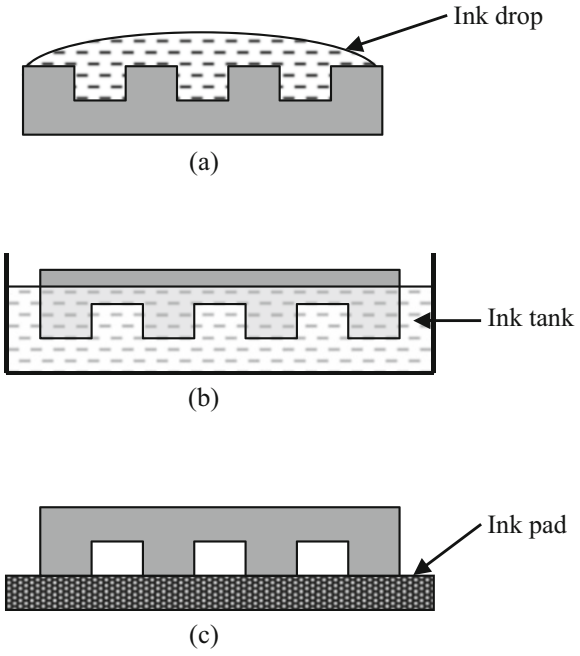


Fig. 6.35 (**a**) Wet inking by ink drop; (**b**) wet inking by dipping into an ink tank; (**c**) dry inking by contact with an ink pad

It is obvious that both of the "wet inking" methods (Fig. 6.35a, b) have thiols homogeneously distributed over all the surface of stamp, including sidewalls and bottom of stamp cavities. When the PDMS stamp is pressed down to print, two things can happen in addition to printing a stamp pattern: one is the diffusion of thiols from cavity sidewalls to the substrate surface, which broaden the printed ink feature; another is the evaporation of thiol-containing solvent from the cavity onto the substrate area where there is not supposed to be an ink pattern. The evaporated traces of thiol can affect the contrast of printed pattern because at a later stage of etching, any traces of thiol can deter the etching of substrate material. The consequence of "wet inking" of stamp is the difficulty to achieve high-resolution printing features. To overcome the above problems, there is a third inking method, as shown in Fig. 6.35c. It uses an ink pad which is a piece of flat PDMS soaked up with thiol solution. The inking of stamp is done by direct contact of stamp with the inking pad. The amount of ink transferred onto the stamp surface is controlled by the concentration of thiol in the solvent. In this way, only the pattern protrusion surfaces have inks. The printed ink pattern will be exactly the same dimension as the pattern structures on the stamp. With the contact-inking method, 100 nm gold line features have been produced [110]. One thing should be noticed that a PDMS stamp with "wet inking" can be used repeatedly without further inking (~ 10 times per inking) because the absorbed ink in PDMS stamp can slowly diffuse out, whereas stamps inked by direct contact can only be used once per inking.

As mentioned earlier in this section, μCP cannot work without the alkanethiol SAM system. Two aspects of the alkanethiol SAM system make it unique in μCP: one is that the thiol molecules can self-assemble into a monolayer on the surface of coinage metals such as Au, Ag, Cu, and Pd. The monolayer-forming process is fast and takes place at room temperature. This is important for making a uniform film with low defect density. Although PDMS stamp is elastic and can form an intimate contact with substrate surface, it can never be guaranteed that the ink can be transferred onto substrate uniformly. Particularly, if the ink molecules congregate into clusters once on another material surface, pinhole type of defects can be easily formed, rendering the ink pattern useless as an etching mask. The self-assembling of thiol molecules on a gold surface is fast and automatic, only needing the inked stamp in contact with the gold surface for a short time between 10 seconds and a minute [99]. Longer contact time between an inked stamp and substrate can further increase the density and coverage of SAM, but it also can induce diffusion of thiol molecules outside the boundary of stamp patterns, making the printed feature larger than intended. The diffusion limits the resolution of μCP technique to above 100 nm. Special ink such as immobilized enzyme can be used, which transfers the stamp pattern by catalytic reaction when in contact with acid coated surface. With enzymatic ink, the resolution of μCP is only limited by the range of motion of the immobilized catalyst, which is about 2 nm [111].

In addition to the self-assembling property, another unique property of thiol-based SAM is its very high resistance to wet chemical etching. For example, gold surface protected by a monolayer of hexadecanethiol (HDT) is 10^6 times slower etched in cyanide (gold etchant) [112]. The two unique properties make the thiol

SAM a perfect masking material. Unfortunately, the thiol SAM system only works with coinage metals; among them, thin film gold is the most used material. Nevertheless, one can always use the patterned gold film as a mask to further transfer into other materials. For example, μCP has been used to pattern PMMA resist layer, in which a 20 nm Au film was coated on top of PMMA resist layer. After contact printing by a PDMS stamp, the HDT inked pattern was transferred into the Au film. Then the Au film was used as etch mask to pattern the PMMA resist layer by oxygen plasma etching [113]. The same concept can be applied to many other multilayer systems as long as the Au or Ag film can act as a masking material in subsequent pattern transfer processes.

6.6.3 Replication by Capillary Force

The unique property of PDMS also enables some other ways of replicating micro- and nanostructures. One of the earliest reported, apart from the μCP, was micromolding in capillaries (MIMIC) [114]. The principle of MIMIC is quite simple, as schematically shown in Fig. 6.36. A PDMS stamp is brought to intimate contact with a flat substrate. A low viscosity liquid polymer is fed into the channels formed by cavities in the stamp. By capillary action, the liquid polymer fills in these

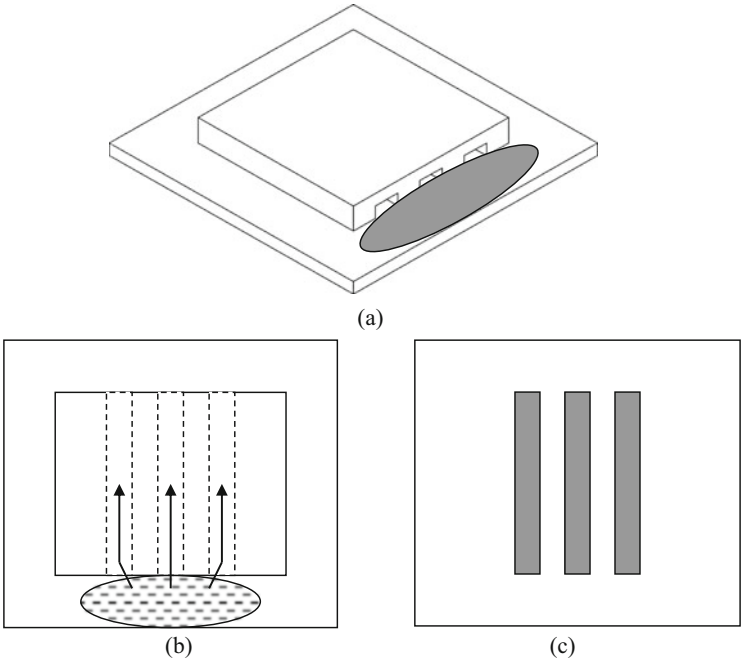


Fig. 6.36 Schematic of micromolding in capillary (MIMIC)

channel cavities (Fig. 6.36a, b). A UV curing or thermal curing solidifies the polymer inside the cavities. After peeling off the PDMS stamp, a polymer replication of the channel pattern is left on the substrate (Fig. 6.36c). Instead of feeding liquid polymer from the side of stamp, the polymer can also be coated on a flat substrate and is heated above its *Tg*, while a PDMS stamp seals from the top. When the polymer lowers its viscosity at high temperature, it can also fill up the stamp cavities with capillary action [115]. Alternatively, the PDMS stamp can be directly placed onto a freshly spin-coated polymer film while it is still wet. Liquid polymer gets into the stamp cavities by capillary force, if the PDMS surface is properly treated and can be wetted. The PDMS stamp can absorb all the solvent contained in the polymer film because of the solvent permeability of PDMS. For example, PDMS can absorb PGMEA (propylene glycol mono ether acetate) up to 27 wt% of its weight, and 30 wt% of absorbed PGMEA will diffuse into air in 1 hour, which creates a partial vacuum inside the PDMS cavity and helps to draw in liquid polymer [116]. Other examples of filling mold cavity by capillary action are the J-FIL and SCIL techniques which have been described previously. Analysis has shown that the capillary pressure is greater than 1 bar when the stamp feature size is smaller than 440 nm. The smaller the stamp feature, the higher the capillary pressure becomes [117].

Another form of replication using capillary force is the solvent-assisted microcontact molding (SAMIM) [118]. The SAMIM process is illustrated in Fig. 6.37. A polymer film is coated onto a substrate and baked. A patterned PDMS stamp is wetted with a matching solvent for the polymer and then brought into contact with the polymer surface. The solvent on the stamp dissolves the polymer, and this dissolved polymer is molded into the channels or cavities formed between the stamp and the substrate by capillary action. The matching solvent for PMMA is acetone, and for photoresists, it is ethanol. The molded polymer structure becomes solidified in a few minutes after evaporation of solvent, while the stamp is still in conformable contact with the substrate. There are two problems associated with the SAMIM technique. The first is that the surface of the patterned polymer is highly textured because of kinetic movement of polymer molecules during solvent dissolving process. The other is that the pattern formed by SAMIM gets distorted. In particular, the base of a rectangular stripe becomes wider than the top.

PDMS is not only permeable to solvent but also permeable to gas. The gas permeability coefficient of PDMS is higher than other polymers. Table 6.1 compares permeability coefficients of PDMS with two other common polymer materials for some typical gases.

It is apparent that PDMS has very high air ($N_2 + O_2$) permeability. This property has been utilized to replicate polymer pattern structure through PDMS molds in a so-called micro-aspiration assisted lithography (MAAL) technique [120]. The MAAL is illustrated in Fig. 6.38. It uses a double-layer mold structure. The top layer consists of thick PDMS (5 mm) with a matrix of channels leading to an outlet which is connected to a vacuum pump. The bottom layer is the actual PDMS mold consisting of various pattern cavities and is separated from the top layer by a thin PDMS membrane (20 μm). Here is the working principle of MAAL: when the composite mold is brought in contact with the polymer-coated flat substrate

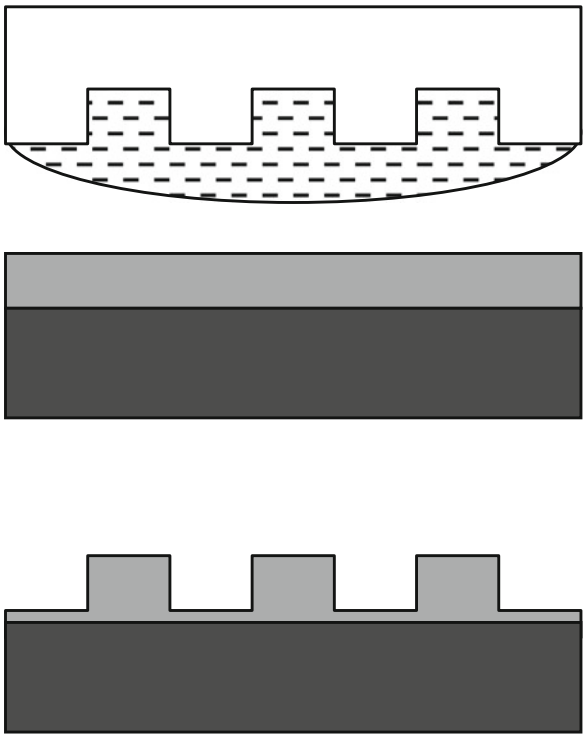


Fig. 6.37 Solvent-assisted microcontact molding (SAMIM)

Table 6.1 Gas permeability coefficients of PDMS, PS, and PMMS

	Permeability coefficient (Barrer)		
Polymers	He	N_2	O_2
PDMS	590	351	781
Polystyrene (PS)	22.4	0.52	2.9
PMMA	8.4	0.02	0.14

(1 Barrer $=10^{-10}$ cm^3 (STP) cm cm^{-2} $cmHg^{-1}$ s^{-1}) [119]

(no prebaking of spin-coated polymer), it seals off the cavities with the liquid polymer layer; a low vacuum (870 mbar) is created by evacuating air from the channels in the top PDMS layer; the pressure difference between the sealed mold cavities and the channels helps air permeating ("aspirating") through separation PDMS membrane, resulting in a vacuum in the mold cavities; the liquid polymer is then drawn into the cavities by the ambient atmospheric pressure. After filling the mold cavities, a thermal or UV curing process can solidify the polymer, and a replication of polymer pattern is formed after peeling of the PDMS mold.

The mold filling process is not assisted by capillary action, but by the air permeability of PDMS and the pressure difference created across the PDMS membrane. The time required to fill the mold cavities depends directly on the air

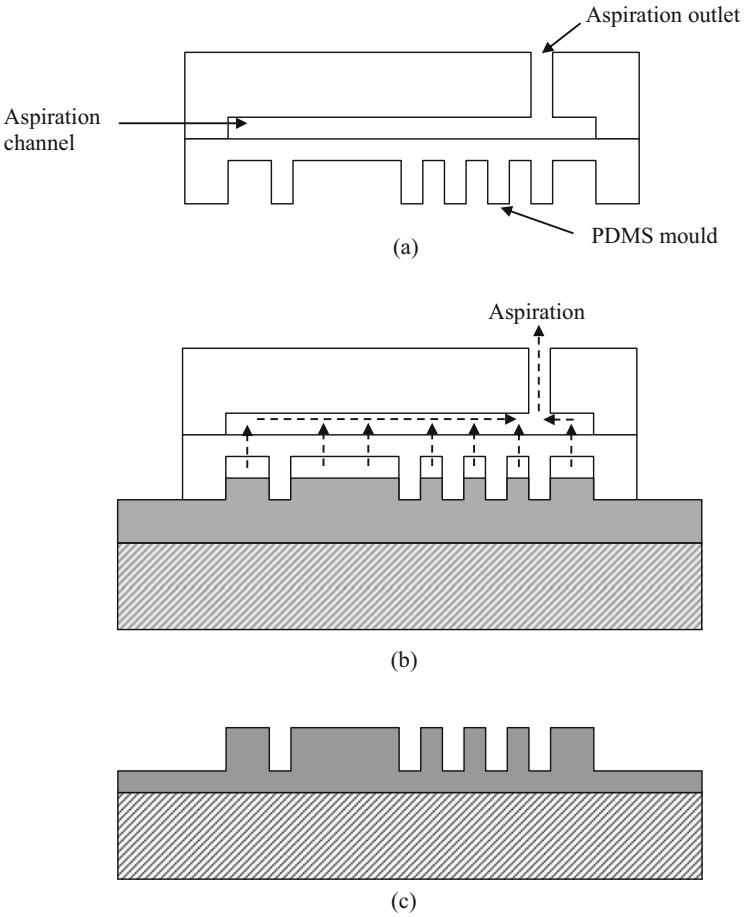


Fig. 6.38 Schematic of micro-aspiration assisted lithography (MAAl) process: (**a**) composite PDMS mold; (**b**) aspiration filling mold vertically; (**c**) aspiration filling mold horizontally. (Reprinted from [120] with permission)

permeability coefficient of PDMS and the pressure difference. If the polymer cannot wet the PDMS mold surface, or the PDMS surface is hydrophobic to the polymer liquid, then the pressure difference has to be greater than the negative capillary force which resists the polymer flowing into the PDMS mold cavity. The polymer filling speed also depends on the viscosity of polymer and geometry of the mold cavities. Nevertheless, MAAL is a much faster process than previously described capillary force lithography. For the same mold area, MAAL only needs 2 minutes to fill up the entire mold cavity while capillary force filling needs 20 minutes, providing the polymer is a wetting material to PDMS [120]. Then for non-wetting polymer, MAAL also works. Despite its advantages, MAAL may have difficulty to fill mold cavity of small geometry if the PDMS surface appears to be hydrophobic to the

filling polymer, because the smaller the cavity geometry, the greater the negative capillary force. If the PDMS surface is hydrophilic to the filling polymer, filling can be accomplished even without the vacuum assistance.

6.7 Roll-to-Roll Continuous Nanoimprint

The replication nature of nanoimprint determines that the technology is useful only if there is a demand for large number of same pattern structures to be made. The patterning is low cost only if the same patterns are replicated many times, because making the mold is expensive. Scientific research does not have many patterns which require replication in large numbers; therefore, it makes no sense to use nanoimprint for research. As mentioned in the Introduction (Sect. 6.1), the technology has found applications in large scale manufacturing of photonic structures in recent years. The photonic structures refer to single layer periodical surface relief patterns at the dimension of sub-100 nm or a few hundreds of nanometers to a few micrometers, which generate optical effects such as antireflection, light guiding or diffusion, light extraction and propagation, holographic imaging, etc. [121]. They usually require large area in the form of flexible thin films, which facilitates roll to roll (R2R) type of continuous nanoimprint [122].

R2R continuous nanoimprint can be done in two different manners—flat nanoimprint and roll nanoimprint—depending on the type of mold. In flat nanoimprint, the mold is a flat plate with surface relief structures. The substrate is flexible plastic film. As illustrated in Fig. 6.39a, the film substrate is rolled in from the unwinding roller, and the mold is pressed down to make the imprint, and then the imprinted film is rolled out by the rewinding roller. It is not a truly continuous process, as the mold can only be pressed in step mode. The process is often called

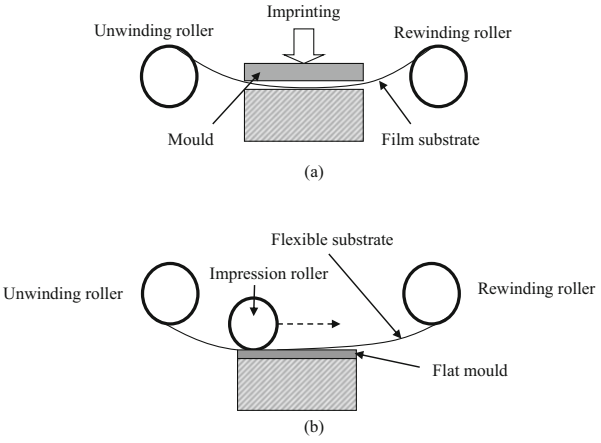


Fig. 6.39 (**a**) Flat nanoimprint in reel to reel fashion; (**b**) roller imprint using a flat mold

reel to reel process. Although the imprint can be done on a long roll of film, the difficulties are how to achieve uniform pressure over the entire imprinted surface each time as the mold is pressed down and how to avoid air bubbles being trapped on the imprinted surface. The solutions include using an airbag to apply the pressure instead of mechanical rigid device and placing the imprint section in a vacuum to ensure no trapped air during imprinting. An option to avoid the above problems is to use an impression roller to press down the film substrate to a flat rigid mold, as shown in Fig. 6.39b. At any moment of imprinting, only a very narrow region of the mold is imprinted. In this way, trapping of air bubbles can be avoided and a uniform pressure can be achieved.

Alternatively, large area flat nanoimprint can be achieved by using a flexible mold pressed by a rigid roller onto substrate. The setup is illustrated in Fig. 6.40. Instead of using a rigid flat mold, surface relief structures are made on a roll of flexible film which serves as the mold. A rigid roller is pressed down on the film mold and rolled over to make the mold imprinting into the underneath substrate which in this case is a rigid substrate. As the imprint is only performed at the area where the rigid roller is in contact with the flat substrate, the contact area is very small, and it is much easier to achieve uniform pressure, as well as much less chance to trap air inside the imprinted area.

The true R2R continuous nanoimprinting is to use a rigid roller mold continuously rolling over a roll of flexible thin film substrate. A simple setup of continuous R2R nanoimprint is illustrated by Fig. 6.41. A photo of real R2R thermal

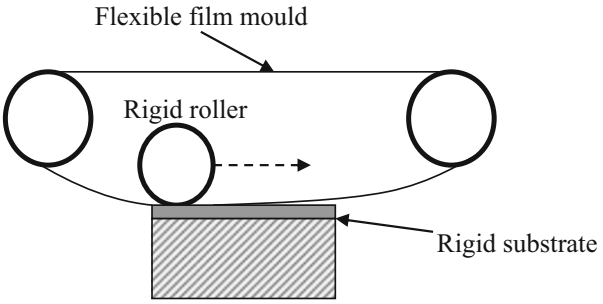


Fig. 6.40 Flat nanoimprint by rolling flexible mold

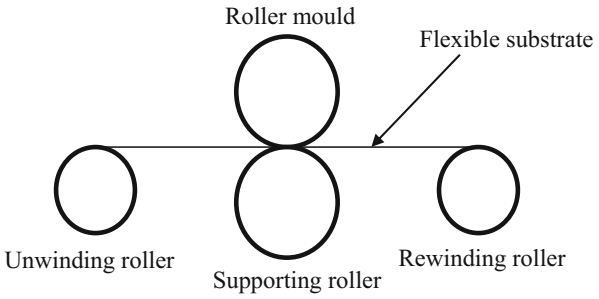


Fig. 6.41 Setup of R2R continuous nanoimprint

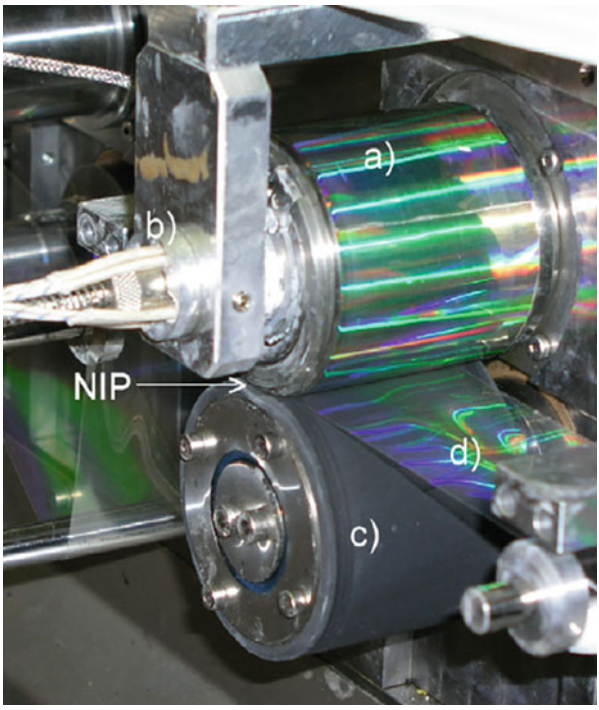


Fig. 6.42 Photo of R2R thermal nanoimprint system: (**a**) roller mold, (**b**) heating device to heat up the roller, (**c**) supporting roller, (**d**) thin film substrate [123]

nanoimprint (T-NIL) system is shown in Fig. 6.42 [123]. Flexible thin film substrate is continuously fed through, while the roller mold is pressed onto the thin film, with the support of supporting roller. The roller mold is equipped with heating device which softens the polymer coated at the surface of film substrate, enabling thermal imprint into the polymer layer. This type of R2R can also be used to imprint on both sides of polymer film if the supporting roller is also a roller mold [30].

As previously mentioned, thermal nanoimprint is not an efficient process because of the time required for heating and cooling. The best option for R2R continuous nanoimprint is UV cured nanoimprint. Typical UV cured R2R nanoimprint is illustrated as in Fig. 6.43. It consists of four basic components: (1) unwinding and rewinding system to provide continuous feeding of film substrate, (2) coating device to coat UV-curable liquid polymer layer onto the thin film substrate, (3) imprinting roller mold which continuously rolls over the coated thin film substrate and impressed into the liquid polymer layer, and (4) UV curing light source which illuminates the thin film substrate from backside, assuming the thin film substrate is transparent, to cure the polymer while the roller mold is still impressed in the polymer layer.

There are a number of challenges in R2R continuous nanoimprint:

1. Making the roller mould. For flat mould, any previously introduced lithographic techniques can be used, such as optical lithography, e-beam lithography and FIB

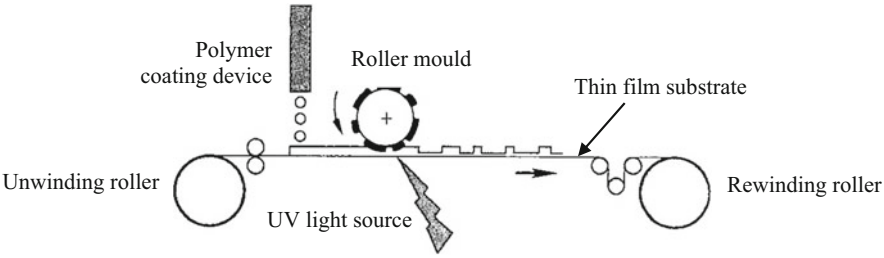


Fig. 6.43 Setup of UV cured R2R continuous nanoimprint

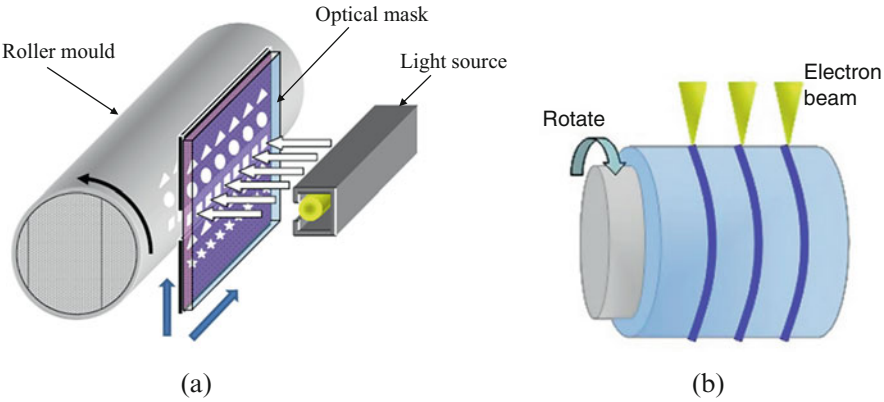


Fig. 6.44 (**a**) Proximity photolithography of roller mold; (**b**) e-beam continuous exposure of roller mold

direct machining. To make continuous patterns on a roller is difficult. For large pattern structures diamond turning tools can be used to cut grooves into the roller surface. The minimum features which can be achieved by diamond turning are limited to above 10μm, unless the diamond tip is made by focused ion beam milling [124]. Besides, the pattern types made by diamond turning on a rotating roller are also limited to certain periodical features. It cannot make arbitrary patterns. A roller mould is usually made on a flexible flat substrate first and then wrapped over the roller, such as the mickle sheet mould shown previously in Fig. 6.4. Although a flat mould sheet is easily to make and can be made with high resolution arbitrary patterns, the drawback is that it is not seamless and there is always a gap when the mould sheet wraps over a roller.

There have been a few attempts to make seamless roller mold with fine features. The reported techniques include proximity photolithography [125], laser direct write [126] and e-beam lithography [127], or self-assembled patterning using anodic aluminum oxide (AAO) [128]. Figure 6.44a illustrates how proximity photolithography is done over the roller surface. A flat photomask is placed within the proximity of a roller. UV light shines through a narrow slit to

illuminate the mask, while the roller rotates in step motion. Line exposure is performed at each rotating step, and the whole surface of roller mold can be exposed. Microlens array has been made onto a roller mold by this method [125]. However, the method cannot make continuous patterns such as long continuous grating because of its step mode. Continuous patterns can be made by continuous exposure on a continuously rotating roller. Such continuous exposures include laser or e-beam direct write, such as shown in Fig. 6.44b. Seamless roller mold with 120 nm line features has been reported by e-beam exposure of rotated roller [129]. For seamless elastic roller mold such as those made in PDMS, it can be made first in the inner surface of a cylinder with photoresist and internal optical lithography and then cast liquid PDMS within the annular space between two cylinders [130].

2. Large area uniform coating of polymer. Nanoimprint, whether on rigid substrates or flexible thin film substrates, requires uniform coating of thermally impressible or UV-curable polymer layer. So far in this book, only one type of coating has been introduced, which is the spin-coating. Spin-coating is only suitable for round shape and rigid substrates such as silicon wafers. Spin-coating on a square substrate will generate nonuniform coating at the four corners. Large area continuous coating of liquid materials is a mature industrial process. There are many types of coating processes that have been applied to NIL, such as rod coating, dip coating, doctor blade coating, air knife coating, slot die coating, gravure coating, screen coating, and inkjet coating [122]. The challenge is to get very thin coating uniformly over a large area. The minimum thickness for conventional coating processes is normally above 1 μm. In recent years, development of organic and printed electronics such as organic light emitting diodes (OLED) and organic photovoltaic (OPV) placed great demands on thin layer coating with precisely controlled thickness. Significant progresses have been made in reducing the coating thickness. Both slot-die [131] and micro gravure [132] coatings can achieve minimum 20 nm thickness.
3. Dynamic alignment and overlay with high accuracy. If more than one layer of patterns is to be made on substrate, R2R continuous nanoimprint will require dynamic alignment and overlay, which is much hard than static alignment and overlay. The usual method is to capture the alignment marks on a continuous rolling substrate with a camera. Any deviations of substrate from roller mold will feedback to winding system to adjust the rolling speed and to roller mold control to finely adjust its lateral position. This can place great demands on the system control, which is not easy to achieve. Presently, the best dynamic overlay accuracy on continuous R2R nanoimprint is around 10 μm, which has severely restricted the overall patterning accuracy of nanoimprint.

Due to the aforementioned challenges, nanoimprint in R2R continuous mode is hardly in nanoscale regime. Although R2R imprinted 100 nm features have been reported in laboratory [133], such patterning dimension has not been realized at industrial scale. Fortunately, most photonic applications do not require very small

feature dimensions and do not require overlay alignment, for which the current R2R continuous imprint processes and technologies have been adequate and played important roles [122].

References

1. Chou, S.Y., Krauss, P.R., Renstrom, P.J.: Imprint of sub-25nm vias and trenches in polymers. Appl. Phys. Lett. **67**(21), 3114–3116 (1995)
2. Zhou, W.: Nanoimprint Lithography: An Enabling Process for Nanofabrication. Springer, Berlin/Heidelberg (2013)
3. Miyauchi, A. (ed.): Nanoimprinting and Its Applications. Jenny Stanford Publishing, Singapore (2019)
4. International Technology Roadmap for Semiconductors. http://www.itrs2.net/itrs-reports.html
5. Chou, S.Y.: Patterned magnetic nanostructures and quantized magnetic disks. Proc. IEEE. **85**(4), 652–671 (1997)
6. Wu, W., et al.: Large area high density quantized magnetic disks fabricated using nanoimprint lithography. J. Vac. Sci. Technol. **B16**, 3825 (1998)
7. Pepin, A., et al.: Nanoimprint lithography for the fabrication of DNA electrophoresis chip. Microelectron. Eng. **61–62**, 927 (2002)
8. Yu, Z., Schablitsky, S.J., Chou, S.Y.: Nanoscale GaAs metal–semiconductor–metal photodetectors fabricated using nanoimprint lithography. Appl. Phys. Lett. **74**, 2381 (1999)
9. Wang, J., Schablitsky, S.J., Chou, S.Y.: Fabrication of a new broadband waveguide polarizer with a double-layer 190 nm period metal-gratings using nanoimprint lithography. J. Vac. Sci. Technol. **B17**, 2957 (1999)
10. Martini, I., et al.: Quantum point contacts fabricated by nanoimprint lithography. Appl. Phys. Lett. **77**, 2237 (2000)
11. Chou, S.Y., et al.: Sub-10 nm imprint lithography and applications. J. Vac. Sci. Technol. **B15**(6), 2897 (1997)
12. Colburn, M., et al.: Step and flash imprint lithography: a new approach to high-resolution patterning. Proc. SPIE. **3676**, 379 (1999)
13. Hamay, Z., et al.: High volume semiconductor manufacturing using nanoimprint lithography. Proc. SPIE. **10810**, 108100F (2018)
14. Ganping, J., et al.: High density heat-assisted magnetic recording media and advanced characterization – progresses and challenges. IEEE Trans. Magn. **51**(11), 3201709 (2015)
15. Kryder, M.: Fifty yeras of disk drives and the exciting road ahead. In: IDEMA DISKCON USA, Seagate Technology, Santa Clara (2006)
16. Malloy, M., Litt, L.C.: Technology review and assessment of nanoimprint lithography for semiconductor and patterned media manufacturing. J. Micro/Nanolith. MEMS MOEMS. **10**(3), 032001 (2011)
17. Albrecht, T.R., et al.: Bit-patterned magnetic recording: theory, media fabrication, and recording performance. IEEE Trans. Magn. **51**(5), 0800342 (2015)
18. Haley, G.: Nanoimprint finally finds its footing. Semicond. Eng. April 20th, 2023
19. Heidari, B., et al.: Large scale nanolithography using nanoimprint lithography. J. Vac. Sci. Technol. **B17**(6), 2961 (1999)
20. Unno, N., Mäkelä, T., Taniguchi, J.: Thermal roll-to-roll imprinted nanogratings on plastic film. J. Vac. Sci. Technol. **32**(6), 06FG03-1 (2014)
21. Haffner, M., et al.: Simple high resolution nanoimprint-lithography. Microelectron. Eng. **84**, 937–939 (2007)
22. Bird, R.B., Amstrong, R.C., Hassager, O.: Fluid mechanics. In: Dynamics of Polymeric Liquids. John Wiley & Sons, New York (1977)

23. Schift, H., Heyderman, L.: Nanorheology: squeeze flow in hot embossing of thin films. In: Torres, C.M.S. (ed.) Alternative Lithography. Kluwer Academic, New York (2003)
24. Halary, J.L., et al.: Viscoelastic properties of styrene-co-methyl methacrylate random copolymers. J. Polym. Sci. B Polym. Phys. **29**(8), 933 (1991)
25. Hoffmann, T.: Viscoelastic properties of polymers: relevance for hot embossing lithography. In: Torres, C.M.S. (ed.) Alternative Lithography. Kluwer Academic, New York (2003)
26. Scheer, H.C., et al.: Problems of the nanoimprinting technique for nanometer scale pattern definition. J. Vac. Sci. Technol. **B16**(6), 3917 (1998)
27. Landis, S., et al.: Stamp design effect on 100 nm feature size for 8 inch NanoImprint lithography. Nanotechnology. **17**, 2701–2709 (2006)
28. Cui, B., Veres, T.: Pattern replication of 100 nm to millimeter-scale features by thermal nanoimprint lithography. Microelectron. Eng. **83**, 902–905 (2006)
29. Schulz, H., et al.: Impact of molecular weight of polymers and shear rate effects for nanoimprint lithography. Microelectron. Eng. **83**, 259–280 (2006)
30. Unno, N., Mäkelä, T.: Thermal nanoimprint lithography – a review of the process, mold fabrication, and material. Nano. **13**, 2031 (2023)
31. Nanonex Corp. http://www.nanonex.com/
32. Microresist GmbH. http://www.microresist.de/
33. Bogdanski, N., et al.: Structure size dependent recovery of thin polystyrene layers in thermal imprint lithography. Microelectron. Eng. **84**, 860–863 (2007)
34. Workum, K.V., de Pablo, J.J.: Computer simulation of the mechanical properties of amorphous polymer nanostructures. Nano Lett. **3**(10), 1405 (2003)
35. Ro, H., et al.: Evidence for internal stresses induced by nanoimprint lithography. J. Vac. Sci. Technol. **B24**(6), 2973 (2006)
36. Ding, Y., et al.: Relaxation behavior of polymer structures fabricated by nanoimprint lithography. ACS Nano. **1**(2), 84–92 (2007)
37. Papenheim, M., Dhim, K., Wang, S.: Effect of residual stress on replication fidelity with nanoimprint. J. Vac. Sci. Technol. **B32**(6), 06FG01-1 (2014)
38. Cox, L.M., et al.: Nanoimprint lithography: emergent materials and methods of actuation. Nano Today. **31**, 100838 (2020)
39. Mekaru, H.: Formation of metal nanostructures by high-temperature imprinting. Microsyst. Technol. **20**, 1103–1109 (2014)
40. Ramachandran, S., et al.: Deposition and patterning of diamondlike carbon as antiwear nanoimprint templates. J. Vac. Sci. Technol. **B24**(6), 2993 (2006)
41. Chou, S.Y., Krauss, P.R., Renstrom, P.J.: Nanoimprint lithography. J. Vac. Sci. Technol. **B14**(6), 4129 (1996)
42. Padeste, C., et al.: Anti-sticking layers for nickel-based nanoreplication tools. Microelectron. Eng. **123**, 23–27 (2014)
43. Shiotsu, T., Nishikura, N., Yasuda, M.: Simulation study on the template release mechanism and damage estimation for various release methods in nanoimprint lithography. J. Vac. Sci. Technol. **31**(6), 06FB07-1 (2013)
44. Konishi, T., et al.: Multi-layered resist process in nanoimprint lithography for high aspect ratio pattern. Microelectron. Eng. **83**, 869–872 (2006)
45. Zhang, W., Chou, S.Y.: Multilevel imprinting lithography with submicron alignment over 4 in Si wafers. Appl. Phys. Lett. **79**(6), 845 (2001)
46. Tan, H., et al.: Current status of Nanonex nanoimprint solutions. SPIE. **5374**, 213–221 (2004)
47. Suss MicroTech. Available from: http://www.suss.com/
48. Ahopelto, J., Haatanien, T.: Step and stamp imprint lithography. In: Torres, C.M.S. (ed.) Alternative Lithography. Kluwer Academic, New York (2003)
49. Chen, Y., et al.: A study of pattern placement error by thermal expansions in nanoimprint lithography. J. Microlith. Microfab. Microsys. **5**(1), 206 (2006)

50. Lebib, A., et al.: Room temperature and low pressure nanoimprint lithography. Microelectron. Eng. **61–62**, 371 (2002)
51. Ge, J., et al.: Rapid fabrication of complex nanostructures using room-temperature ultrasonic nanoimprinting. Nat. Commun. **12**, 3146 (2021)
52. Yutaka, I., et al.: Room temperature nanoimprint technology using hydrogen Silsequioxane (HSQ). Jpn. J. Appl. Phys. **41**, 4198–4202 (2002)
53. Nakamatsu, K., Matsui, S.: Room-temperature nanoimprint and nanocontact technologies. In: Busnaina, A. (ed.) Nanomanufacturing Handbook. CRC Press, Boca Raton (2007)
54. Tao, J., et al.: Room temperature nanoimprint lithography using a bilayer of HSQ/PMMA resist stack. Microelectron. Eng. **78–79**, 665–669 (2005)
55. Matsui, S., et al.: Room temperature replication in spin on glass by nanoimprint technology. J. Vac. Sci. Technol. **B19**(6), 2801 (2001)
56. Lu, Y., et al.: Patterning layered polymeric multilayer films by room-temperature nanoimprint lithography. Macromol. Rapid Commun. **27**(7), 505–510 (2006)
57. Haisma, J., et al.: Mold-assisted nanolithography: a process for reliable pattern replication. J. Vac. Sci. Technol. **B14**, 4124 (1996)
58. Voisin, P., et al.: High-resolution fused silica mold fabrication for UV-nanoimprint. Microelectron. Eng. **84**, 916–920 (2007)
59. Dauksher, W.J., et al.: Characterization of and imprint results using indium tin oxide-based step and flash imprint lithography templates. J. Vac. Sci. Technol. B. **20**, 2857–2861 (2002)
60. Intermediate Polymer Stamp (IPS™). https://www.obducat.com/
61. OrmoStamp®. https://www.microresist.de/
62. Kawaguchi, Y., Nonaka, F., Sanada, Y.: Fluorinated materials for UV nanoimprint lithography. Microelectron. Eng. **84**, 973–976 (2007)
63. Otsuka, Y., Hiwasa, S., Taniguchi, J.: Development of release agent-free replica mould material for ultraviolet nanoimprinting. Microelectron. Eng. **123**, 192–196 (2014)
64. Asif, M.H., et al.: Comparison of UV-curable materials for high-resolution polymer nanoimprint stamps. Micro Nano Eng. **14**, 100118 (2022)
65. Cheng, X., Guo, L.J.: One-step lithography for various size patterns with a hybrid mask-mold. Microelectron. Eng. **71**, 288–293 (2004)
66. Guo, L.J.: Recent progress in nanoimprint technology and its applications. J. Phys. D. Appl. Phys. **37**, R123–R141 (2004)
67. Wang, X., et al.: High density patterns fabricated in SU-8 by UV curing nanoimprint. Microelectron. Eng. **84**, 872–876 (2007)
68. Peroz, C., et al. :Nanoimprint Technologies. in Nanofabrications Techniques and Principles, Maria Stepanova and Steven Dew (eds), Springer WienNewYork (2012)
69. Bender, M., et al.: Multiple imprinting in UV-based nanoimprint lithography: related material issues. Microelectron. Eng. **61–62**, 407–413 (2002)
70. Vogler, M., et al.: Development of a novel, low-viscosity UV-curable polymer system for UV-nanoimprint lithography. Microelectron. Eng. **84**, 984–988 (2007)
71. Min, H., et al.: UV-curable nanoimprint resist with liquid volume-expanding monomers. Microelectron. Eng. **205**, 32–36 (2019)
72. Voisin, P., et al.: Characterisation of ultraviolet nanoimprint dedicated resists. Microelectron. Eng. **84**, 967–972 (2007)
73. Guo, L.J.: Nanoimprint lithography: methods and material requirements. Adv. Mater. **19**, 495–513 (2007)
74. Le, N.V., et al.: Selective dry etch process for step and flash imprint lithography. Microelectron. Eng. **78–79**, 464–473 (2005)
75. Thanner, C., Eibelhuber, M.: UV nanoimprint lithography: geometrical impact on filling properties of nanoscale patterns. Nano. **11**, 822 (2021)
76. Gerard, S., et al.: Jet and flash imprint lithography for the fabrication of patterned media drives. Proc. SPIE. **7488**, 748820 (2009)

77. Sreenivasan, S.V.: Nanoimprint lithography steppers for volume fabrication of leading-edge semiconductor integrated circuits. Microsyst. Nanoeng. **3**, 17075 (2017)
78. Traub, M.C., Longsine, W., Truskett, V.N.: Advances in nanoimprint lithography. Annu. Rev. Chem. Biomol. Eng. **7**, 583–604 (2016)
79. Resnick, D.J., Sreenivasan, S.V., Willson, C.G.: Step & flash imprint lithography. Mater. Today. **8**(2), 34–42 (2005)
80. Asano, T., et al.: The advantages of nanoimprint lithography for semiconductor device manufacturing. Proc. SPIE. **11178**, 111780 (2019)
81. Ji, R., et al.: UV enhanced substrate conformal imprint lithography (UV-SCIL) technique for photonic crystals patterning in LED manufacturing. Microelectron. Eng. **87**, 963–967 (2010)
82. Verschuuren, M.A., et al.: Large area nanoimprint by substrate conformal imprint lithography (SCIL). Adv. Opt. Technol. **6**(3–4), 243–264 (2017)
83. SCIL tool kit. https://www.suss.com/
84. Rumle, M., et al.: Large area manufacturing of plasmonic colour filters using substrate conformal imprint lithography. Nano Futures. **1**, 01500 (2017)
85. Choi, B.J., et al.: Layer-to-layer alignment for step and flash imprint lithography. Proc. SPIE. **4343**, 436–442 (2001)
86. Moel, A., et al.: Novel on-axi interferometric alignment method with sub-10nm precision. J. Vac. Sci. Technol. **B11**(6), 2191 (1993)
87. Muhlberger, M., et al.: A Moiré method for high accuracy alignment in nanoimprint lithography. Microelectron. Eng. **84**, 925–927 (2007)
88. Huang, X.D., et al.: Reversal imprinting by transferring polymer from mold to substrate. J. Vac. Sci. Technol. **B 20**, 2872 (2002)
89. Kehagias, N., et al.: Three-dimensional polymer structures fabricated by reversal ultraviolet-curing imprint lithography. J. Vac. Sci. Technol. **B 23**(6), 2954 (2005)
90. Sogo, K., et al.: Reproduction of fine structures by nanocasting lithography. Microelectron. Eng. **84**, 909–911 (2007)
91. Song, J., et al.: Fabrication of ferroelectric polymer nanostructures on flexible substrates by soft-mold reverse nanoimprint lithography. Nanotechnology. **27**, 015302 (2016)
92. Yoshikawa, T., et al.: Fabrication of 1/4 wave plate by nanocasting lithography. J. Vac. Sci. Technol. **B 23**(6), 2939 (2005)
93. Hirai, Y., et al.: Fine pattern transfer by nanocasting lithography. Microelectron. Eng. **78–79**, 641 (2005)
94. Bao, L.R., et al.: Nanoimprinting over topography and multilayers three-dimensional printing. J. Vac. Sci. Technol. **20**(6), 2881 (2002)
95. Han, K.-S., et al.: Fabrication of 3D nano-structures using reverse imprint lithography. Nanotechnology. **24**, 045304 (2013)
96. Kumar, A., Whitesides, G.M.: Features of gold having micrometer to centimeter dimensions can be formed through a combination of stamping with an elastomeric stamp and an alkanethiol "ink" followed by chemical etching. Appl. Phys. Lett. **63**(14), 2002 (1993)
97. Lipomi, D.J., Martinez, R.V., Cademartiri, L., Whitesides, G.M.: Soft lithographic approaches to nanofabrication. In: Matyjaszewski, K., Moller, M. (eds.) Polymer Science: A Comprehensive Reference. Elsevier, Amsterdam (2012)
98. Zhao, X.M., Xia, Y.N., Whitesides, G.M.: Soft lithographic methods for nano-fabrication. J. Mater. Chem. **7**(7), 1069–1074 (1997)
99. Xia, Y., Whitesides, G.M.: Soft lithography. Angew. Chem. Int. Ed. **37**, 550–575 (1998)
100. Weibel, D.B., et al.: Microfabrication meets microbiology. Nat. Rev. Microbiol. **5**, 209–218 (2007)
101. Miranda, I., et al.: Properties and applications of PDMS for biomedical engineering: a review. J. Funct. Biomater. **13**(2), 2–20 (2022)
102. Tormen, M.: Microcontact printing techniques. In: Torres, C.M.S. (ed.) Alternative Lithography. Kluwer Academic, New York (2003)

103. Hui, C.Y., et al.: Constraints on microcontact printing imposed by stamp deformation. Langmuir. **18**, 1394–1407 (2002)
104. Delamarche, E., et al.: Stability of molded polydimethylsiloxane microstructures. Adv. Mater. **9**(9), 741–746 (1997)
105. Schmid, H., Michel, B.: Siloxane polymers for high-resolution, high-accuracy soft lithography. Macromolecules. **33**, 3042–3049 (2000)
106. Tormen, M., et al.: Sub-mm thick rubber-elastic stamp on rigid support for high reliability microcontact printing. Microelectron. Eng. **61–62**, 469–473 (2002)
107. Plachetka, U., et al.: Comparison of multilayer stamp concepts in UV–NIL. Microelectron. Eng. **83**, 944–947 (2006)
108. Hu, X., et al.: High resolution soft mold for UV-curing nanoimprint lithography using an oxygen insensitive degradable material. J. Vac. Sci. Technol. B. **32**(6), 06FG07-1 (2014)
109. Yoo, P.J., et al.: Unconventional patterning with a modulus-tunable mold: from imprinting to microcontact printing. Chem. Mater. **16**, 5000–5005 (2004)
110. Libioulle, L., et al.: Contact-inking stamps for microcontact printing of Alkanethiols on Gold. Langmuir. **15**, 300–304 (1999)
111. Snyder, P.W., et al.: Biocatalytic microcontact printing. J. Org. Chem. **72**, 7459–7461 (2007)
112. Biebuyck, H.A., et al.: Lithography beyond light: microcontact printing with monolayer resists. IBM J. Res. Dev. **41**(1/2), 159 (1997)
113. Chen, Y., et al.: Microcontact printing and pattern transfer with a tri-layer processing. Microelectron. Eng. **53**, 253–256 (2000)
114. Kim, E., Xia, Y., Whitesides, G.M.: Polymer microstructures formed by moulding in capillaries. Nature. **376**, 581 (1995)
115. Suh, K.Y., Lee, H.H.: Capillary force lithography: large-area patterning, self-organisation and anisotropic deweting. Adv. Funct. Mater. **12**(6–7), 406 (2002)
116. Kim, Y.S., Suh, K.Y., Lee, H.H.: Fabrication of three-dimensional microstructures by soft molding. Appl. Phys. Lett. **79**(14), 2285 (2001)
117. Yoon, H., et al.: Capillary force lithography with impermeable molds. Appl. Phys. Lett. **88**, 254104 (2006)
118. Kim, E., et al.: Solvent-assisted microcontact molding: a convenient method for fabricating three-dimensional structures on surfaces of polymers. Adv. Mater. **9**, 651 (1997)
119. Eddings, M.A., Gale, B.K.: A PDMS-based gas permeation pump for on-chip fluid handling in microfluidic devices. J. Micromech. Microeng. **16**, 2396–2402 (2006)
120. Berre, M.L., et al.: Micro-aspiration assisted lithography. Microelectron. Eng. **84**, 864–867 (2007)
121. Lova, P., Soci, C.: Nanoimprint lithography: toward functional photonic crystals. In: Comoretto, D. (ed.) Organic and Hybrid Photonic Crystals. Springer, Cham (2015)
122. Peng, Z., et al.: Continuous roller nanoimprinting: next generation lithography. Nanoscale. **15**, 11403 (2023)
123. Mäkelä, T., Haatainen, T., Ahopelto, J.: Roll-to-roll printed gratings in cellulose acetate web using novel nanoimprinting device. Microelectron. Eng. **88**, 2045–2047 (2011)
124. Sun, J.: Fabrication of periodic nanostructures by single-point diamond turning with focused ion beam built tool tips. J. Micromech. Microeng. **22**, 115014 (2012)
125. Li, W.-H., et al.: Fabrication of seamless roller molds using step and rotate curved surface photolithography and application on micro-lens array optic film. In: Proceedings of the 2011 6th IEEE International Conference on Nano/Micro Engineered and Molecular Systems, pp. 728–731. IEEE, Kaohsiung (2011)
126. Lee, Y.-C., Chen, P.-C., Lin, H.-Y.: Fabrication of seamless roller mold with excimer laser direct writing technology. In: Proceedings of the 2009 6th IEEE International Conference on Nano/Micro Engineered and Molecular Systems, pp. 767–770. IEEE, Shenzhen (2009)
127. Unno, N., Taniguchi, J., Ishikawa, K.: Fabrication of a seamless roll mold using inorganic electron beam resist with postexposure bake. J. Vac. Sci. Technol. B. **29**(6), 06FC06-1 (2011)

128. Yanagishita, T., et al.: Optimization of antireflection structures of polymer based on nanoimprinting using anodic porous alumina. J. Vac. Sci. Technol. B. **26**, 1856–1859 (2008)
129. Unno, N., Taniguchi, J.: Fabrication of the metal nano pattern on plastic substrate using roll nanoimprint. Microelectron. Eng. **88**, 2149–2153 (2011)
130. Tsai, S.-W., et al.: Fabrication of seamless roller mold with 3D micropatterns using inner curved surface photolithography. Microelectron. Eng. **150**, 19–25 (2016)
131. Slot die coating. https://ntact.com/applications/slot-die-coating/
132. Hwang, J.K., et al.: A development and evaluation of micro-gravure coater for printed electronics. Jpn. J. Appl. Phys. **53/56**, 05HC12 (2014)
133. Ahn, S.H., Guo, L.J.: High-speed roll-to-roll nanoimprint lithography on flexible plastic substrates. Adv. Mater. **20**, 2044–2049 (2008)

Chapter 7
Nanoscale Pattern Transfer by Etching



7.1 Introduction

The lithography process is not finished until a functional pattern structure has been made on or into a substrate material. There are of course many ways of making a pattern directly out of a material. Some of them have already been introduced in previous chapters. For example, focused ion beam can directly sputter etching a nanoscale pattern into a substrate material, or make a structure on top of a substrate by gas assisted deposition of a material, as described in Chap. 4. Electron beam, though not able to sputter etching, can also deposit materials with gases, as described in Chap. 3. Both of the techniques do not need a mask, because the charged particle beams can be focused and scanned into any patterns. The same is true for patterning by a scanning probe. Etching or deposition of materials can be done by a probe interacting with a substrate, which has been explained in detail in Chap. 5. However, the majority of nanofabrications involve lithographic patterning of polymer resists. Generally, the polymeric patterns have no other functions except serving as a mask. The patterns in the mask have to be transferred to underlying layers to form functional structures.

There are many ways of transferring a polymer resist pattern into a substrate. These processing techniques can be grouped into either subtractive processes or additive subtractive processes, i.e., removing underlying material from the openings of polymer resist mask or adding a new material through the openings of polymer resist mask. This chapter will introduce the subtractive pattern transfer techniques. The additive pattern transfer will be introduced in Chap. 8. Figure 7.1 illustrates the subtractive pattern transfer processes. Figure 7.1a is pattern transferring to a thin film deposited on a substrate by etching through the resist mask opening, whereas Fig. 7.1b shows pattern transferring directly into a substrate by etching through the resist mask openings.

Subtractive pattern transfer is mostly characterized by various etching techniques, whether wet or dry etching. In wet etching processes, the etchants are liquid form of

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_7

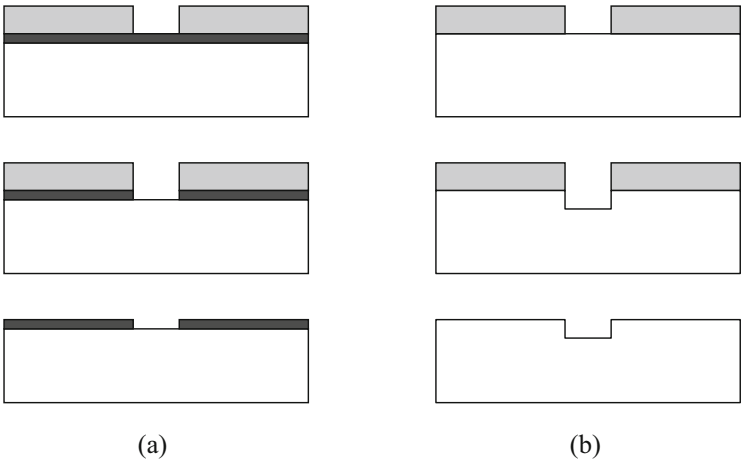


Fig. 7.1 Schematic of subtractive pattern transfer processes: (**a**) etching into a thin film; (**b**) etching into substrate

chemicals. In dry etching, the etchants are gas form of chemicals or in some cases are energized ions. Lithography and etching are the two most important processes which go hand in hand in most micro- and nanofabrication processes. The main character of IC manufacturing is the repeated processes of lithography and etching, which leads to the buildup of multilevel circuit pattern layouts in silicon. Making any other micro- and nanostructures, as long as processing by planar technology, cannot escape one form or another of lithography and etching processes. This chapter will only focus on the etching techniques that enable pattern transfer at sub-100 nm dimension.

Regarding etching, no matter wet or dry, there are two key parameters which characterize all the etching techniques.

(i) Selectivity. Etching is a process of removing substrate material or thin film material on a substrate through mask openings. The mask can be a polymer resist or another thin film material such as silicon dioxide and silicon nitride. The thin film mask is also made by lithography and etching. Selectivity is the ratio of etching rate for substrate or thin film over-etching rate for the mask. High selectivity means there is little loss of masking material and the mask can endure long etching in an etchant.

(ii) Anisotropy. The purpose of etching is to transfer the pattern in mask into the material underneath the mask. Ideally, etching should transfer the exact dimension of mask pattern. In practice, there are always distortions, or lateral etching happens. The lateral etching can be caused by excessive etching of masking layer, which is to do with the aforementioned selectivity issue; the lateral etching can also be the result of lateral etching of the underneath material whether it is the substrate or the thin film on a substrate. The degree of lateral etching is characterized by anisotropy. Figure 7.2 illustrates three different

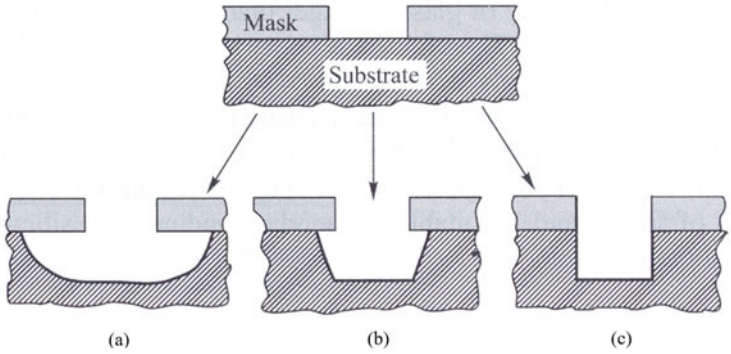


Fig. 7.2 Schematic of anisotropy of etching: (**a**) isotropic, (**b**) partially anisotropic, (**c**) completely anisotropic

anisotropies in etching. It ranges from completely non-anisotropic or called isotropic etching to partially anisotropic etching and to completely anisotropic etching. In general, anisotropic etching transfers mask pattern in its exact dimension into a substrate. However, there are exceptions that anisotropic etching may not follow the mask pattern boundary, such as the alkaline-based wet chemical etching of crystalline silicon. On the other hand, isotropic etching may still be able to transfer the exact dimension of a mask pattern, if the etching depth is extremely shallow.

Ideally, a mask pattern should be faithfully transferred into another material without any distortions. In practice, this is a difficult task. Apart from aforementioned anisotropy issue, each pattern transfer technique has its own capability and limitation, particularly when dealing with sub-100 nm pattern transfer. Some of the pattern transfer techniques are inherently not able to replicate the dimension of mask patterns. Even those techniques which are capable of nanoscale pattern transfer are still highly dependent on various process conditions to achieve desired accuracy and structure profile. The principles of these techniques are described here, and particular attentions are drawn to the conditions at which nanoscale pattern transfer becomes possible.

7.2 Wet Chemical Etching

7.2.1 Wet Isotropic Etching

Wet chemical etching is generally isotropic, i.e., etching progresses in depth as well as in lateral direction. Even though there is a mask, liquid etchant can penetrate underneath the mask, resulting in the etched pattern larger than the mask pattern, as illustrated in Fig. 7.2a. For this reason, wet chemical etching is normally not used for

high resolution pattern transfer. A typical example is the fabrication of photomasks. Photomasks for above 0.25 μm IC technology are made by wet acidic etching of chrome layer using a patterned polymer resist as masking layer. For below 0.25 μm IC technology, various resolution enhancement techniques are used, which include phase-shifting masks (PSM) and optical proximity correction (OPC) masks. PSM and OPC masks both contain sub-resolution pattern features which are in the order of 100 nm. Wet chemical etching is no longer capable of transferring these sub-resolution features faithfully into the chrome layer. Reactive ion etching (RIE) has to be used.

However, the etching in lateral direction is only as fast as in the depth direction. If the etch depth is very small, the lateral etching can be negligible. Therefore, high resolution pattern transfer is still possible. Wet chemical etching has indeed found its way into sub-100 nm pattern transfer applications when self-assembled monolayer (SAM) is used as a resist. Not all SAMs are qualified as a resist for nanoscale pattern transfer. The only SAM that has found applications in nanolithography is the family of alkanethiol SAM systems. Patterning of SAM can be done by microcontact printing (μCP), as described in Chap. 6, or by a scanning probe (dip-pen nanolithography) as described in Chap. 5. The layer thickness of a SAM pattern is only 1–2 nm. Transfer of SAM pattern has to be carried out in wet chemical etching for two reasons: one is that the SAM is so thin that it cannot withstand ion bombardment in a RIE process; the other is that there are no effective RIE processes suitable for etching coinage metals. To realize sub-100 nm scale pattern transfer in wet chemical etching, one needs to make the metal film extremely thin. 20 nm lines have been produced in 10-nm-thick gold film using wet chemical etching of the gold film [1]. Although only coinage metals can be patterned using the SAM resist as mask, the pattern can be further transferred into other materials using the patterned metallic film as a mask. For example, various silicon structures have been generated by anisotropic etching of silicon with patterned gold or silver as the etching masks [2].

Wet chemical etching is inherently isotropic and generally carried out in acidic solutions. For isotropic etching of silicon, the most widely used acidic etchant is the mixture of HF, HNO_3, and acetic acid, commonly called HNA etchant. The mechanism of etching silicon by HNA is that HNO_3 first oxidizes silicon and HF then dissolves the silicon dioxide. Pure HF acid has almost no etching on silicon. Acetic acid here functions as a buffer solution and can be replaced by water. The etch rate depends on the concentration ratio of each of the acids and is represented by a three-phase diagram [3]. HNA has been used as standard etchant in semiconductor industry since the 1960s, mostly for cleaning of silicon wafers. High etch rate is not needed in most silicon IC production processes. Silicon dioxide is another important material and can be isotropically etched. The etchant for silicon dioxide etching is mainly HF acid. HF without dilution (49% weight) has etching rate up to 1.8 μm/min. In practice, buffered HF is often used. Buffered oxide etchant (BOE) consists of seven parts of NH_4F and one part of HF (49% weight) and has etch rate of 70 nm/min for LPCVD oxide. The higher proportion of buffer solution, the lower the oxide etch rate.

When carrying out wet etching with polymer resist mask, one should be aware of the residual resist issue. Lithographically patterned and wet developed polymer resist would normally leave a thin layer (a few nanometers' thick) of polymer residues on substrate surface. Polymers which are mainly hydrocarbons are generally resistant to acidic solutions. Even a very thin layer of polymer resist residue (scum) can stop the wet etching of substrate. Therefore, a descum process usually follows the development of resist pattern before proceeding to acidic etching of substrate. Descum is normally done by a short oxygen plasma etching of developed resist pattern, assuming that the residue layer is very thin. For example, of the wet etching pattern transfer in photomask making process, there is always a descum step before the wet etching of chrome layer. The descum etching by oxygen plasma can, however, affect the resist pattern profile. So it has to be brief, usually less than 30 seconds at low plasma density. Another caution is the effect of descum on substrate surface. In principle, the descum etching should not damage the underlying thin film or substrate surface.

7.2.2 Wet Anisotropic Etching

Although wet chemical etching in general is isotropic, an exception is silicon etching by alkaline-based etchant, which is dependent on crystalline orientation; therefore, it is anisotropic. The anisotropy comes from the fact that alkaline dissolves silicon in different rates at different crystalline planes. The most commonly encountered crystalline planes of silicon wafers are <100>, <111>, and < 110>, which are illustrated in Fig. 7.3. The etch rate decreases in the order of <110>, <100>, and < 111> for the reason that they have different density of atoms on these planes. The etch rate in <110> plane is twice higher than the <100> plane, and they both are hundreds of times higher than the <111> plane. Typically, the ratio of etch rates for the three crystalline planes <110>/<100>/<111 > are 400/200/1 [4]. Because different crystalline planes intersect in different angles, the etch profile of silicon in alkaline etchant can be quite complicated. For example, etching of <100> oriented silicon wafer proceeds fast in <100> direction but slow in <111> direction.

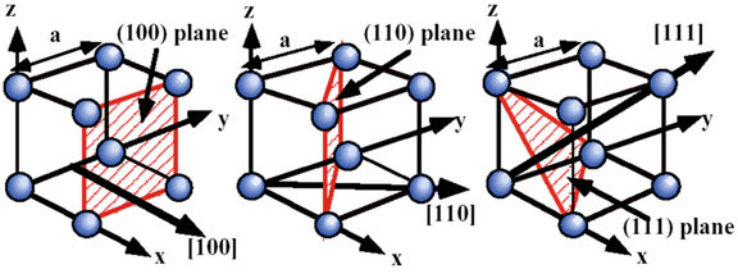


Fig. 7.3 Three main lattice structures of single crystal silicon

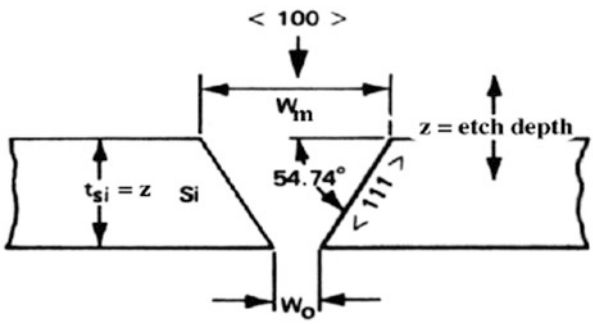


Fig. 7.4 Anisotropic etch of single crystal silicon at <100> plane

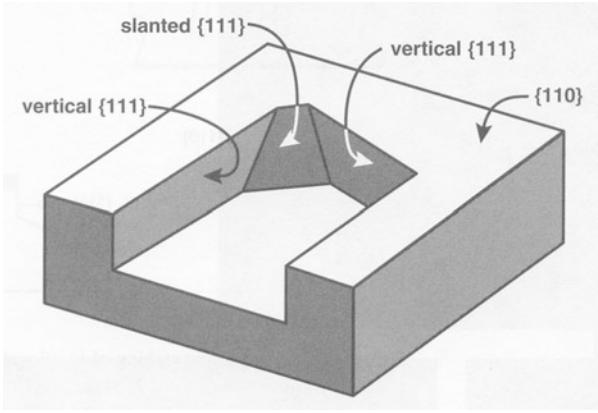


Fig. 7.5 Anisotropic etch of single crystal silicon at <110> plane

Because the <111> plane intersects <100> plane at an angle of 54.74°, the etched profile is cone-shaped with the slope of 54.74°. For limited thickness of silicon wafer, the eventual hole etched through the silicon wafer will have large opening at the top and small opening at the bottom, as illustrated in Fig. 7.4. The etch depth and opening can be calculated by Eq. 7.1.

$$W_0 = W_m - \sqrt{2}z, \text{where } 2\cot an(54.74^\circ) \cong \sqrt{2} \tag{7.1}$$

Chemical etching in <110> direction is also anisotropic. The etched profile in <110> direction is more complicated because <111> and <110> intersect in different angles from different directions. Figure 7.5 illustrates the profile etched at <110> plane. There are eight <111> planes which intersect with <110> plane. Four of them are perpendicular to <110>, and the other four intersect with <110> plane in 35.26° angle. The eight <111> planes themselves are not perpendicular to each other but with angles. Therefore, the etched structures from <110> plane have some sidewalls vertical but some are not. For the <100> oriented silicon wafer, there are four <111> planes intersecting with <100> plane in 54.74° angle. The four <111> planes themselves are vertical to each other. Etching through a square

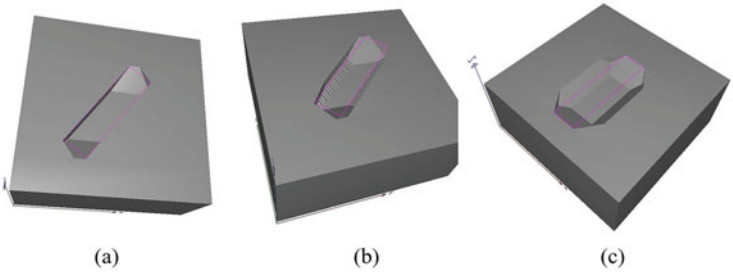


Fig. 7.6 Computer simulated alkaline etching of <110> silicon wafer with a rectangular mask pattern in different orientations: (**a**) perfectly aligned mask, (**b**) slightly misaligned mask, (**c**) largely misaligned mask

or rectangular mask opening will result in a square or rectangular opening in the <100> plane. However, this is not so for <110> plane. A square or rectangular mask opening does not produce a square or rectangular etched hole in <110> oriented silicon wafer.

The dependence on crystal planes highlights a critical issue in wet chemical etching of silicon, i.e., the alignment of etch mask with crystalline axis. Misalignment of mask pattern edge with crystal axis will result in the etched edge following the crystal axis, not the mask pattern edge. Figure 7.6 is a group of simulated examples. The simulation was done using the AnisE software which is a part of IntelliSuite software package [5]. The silicon substrate is <110> oriented. The mask pattern is a rectangular opening. Figure 7.6a, b, c are simulated etched structures for three different alignments of the mask with <110> direction. Only in Fig. 7.6a the mask is perfectly aligned to crystal axis, which results in vertical sidewalls along the mask edge. Figure 7.7 is the experimental proof of such misalignment effect. Figure 7.7a is the etched trench from misaligned mask, where roughness and slanting of sidewall along the trench are apparent. Figure 7.7b is the etched trench when mask is accurately aligned to <110> axis. The sidewall along the trench is smooth and vertical. Recently, a report showed that by aligning the edges of patterned holes to the <111> plane of Si <110> wafer, the KOH wet etch could eliminate the scalloping roughness of sidewall while avoiding measurable lateral etching to the walls [6].

In order to achieve accurate alignment, one must first be aware of the crystalline orientation of a silicon wafer. The commercial silicon wafers come in two crystal planes, <100> plane and < 111> plane. They also have n-type doping and p-type doping. The standard markings of crystal plane and doping type are shown in Fig. 7.8. The <100> oriented wafer means the <100> axis is perpendicular to the wafer surface, whereas the <110> axis is parallel to the primary cut flat. When aligning a mask, one should make sure the edge of square patterns is aligned to the primary cut flat. The marking by wafer vendors is not always accurate. It is better to measure the crystal plane orientation experimentally. For example, one can carry out etching of a series of trenches along different directions and find the trench which

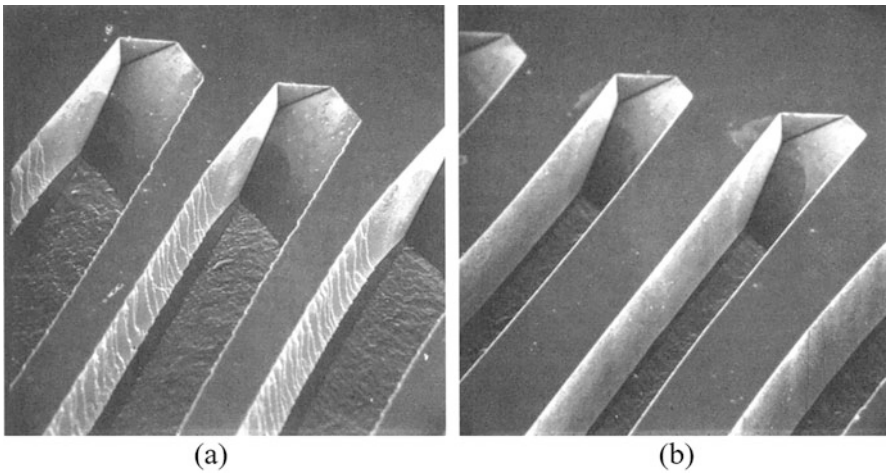


Fig. 7.7 Experimental results of etching in <110> plane: (**a**) slightly misaligned mask, (**b**) perfectly aligned mask

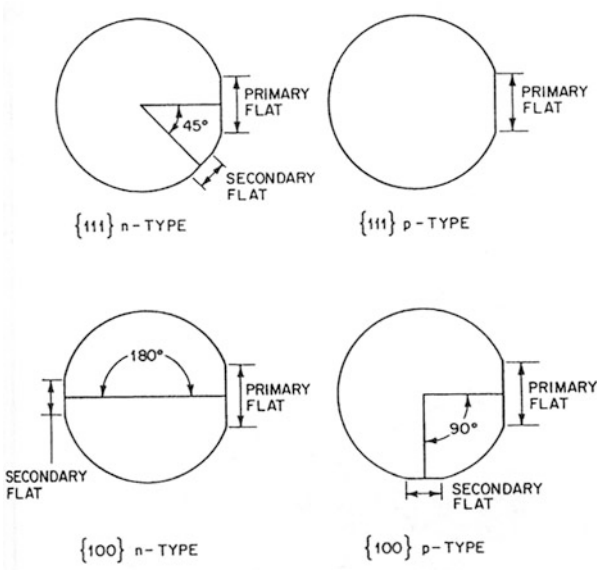


Fig. 7.8 Standard markings of silicon wafer orientation and doping type

best aligned to the crystal axis and use the trench as the mark for mask alignment. This method can minimize the misalignment to less than 0.01° [7].

There are three common alkaline etchants for wet chemical etching of silicon, the KOH (potassium hydroxide), EDP (ethylenediamine pyrocatechol), and TMAH (trimethyl ammonium hydroxide). For etching large structures, all the three etchants are fine. There are some differences in selectivity and etch rate. For example, KOH

has higher silicon etch rate than the other two etchants, but also higher etch rate to oxide. More importantly, KOH etching of silicon creates rough surface, which is highly detrimental when the etched structure itself is at sub-100 nm dimension. EDP has lower oxide etch rate and is less anisotropic because it etches <111> plane faster than KOH does, but it is very toxic and special caution has to be taken when using it. TMAH is the best suitable alkaline etchant for nanoscale etching of silicon because the etched surface is ten times smoother than etching by KOH. 40 nm deep grooves with 50 nm width and 30 nm high ridge with 40 nm width have been fabricated on Si<100> surface with TMAH etchant [8]. The other option is to use a mixture of KOH and IPA (isopropyl alcohol), which can also significantly reduce the surface roughness [9]. The additional benefit of using TMAH is that it does not contain alkaline metals like the KOH. KOH is not compatible with semiconductor fabrication due to the alkaline metal potassium which can threaten transistor gate oxide reliability.

Alkaline etching of silicon is a bulk silicon micromachining process which is generally used to create large structures of tens or hundreds of micrometers. With a < 110> silicon wafer, the technique can be employed for nanoscale pattern transfer as long as the pattern edges are aligned to certain crystalline planes and the etch depth is limited. Line patterns less than 40 nm have been transferred into silicon by alkaline wet etching on <110> silicon wafer, which creates vertical sidewall profile with more than 5:1 aspect ratio and nearly atomic smooth sidewall surface [10]. In conventional bulk micromachining of silicon, large quantity of silicon has to be etched away. A good masking material with high selectivity is necessary. Silicon nitride is preferred because it has higher etching resistance in alkaline etchant than silicon dioxide. For nanoscale wet etching of silicon, the selectivity of masking material is less of an issue because the etch depth is so small. Oxide mask is generally used because an oxide mask pattern can be directly generated by local oxidation using an AFM scanning probe [11]. Otherwise, an oxide mask pattern is generated by electron beam lithography of PMMA resist and subsequent isotropic etching by buffered HF or conventional reactive ion etching process [9]. Only a very thin oxide masking layer is needed for such nanoscale etching.

7.2.3 Metal Assisted Anisotropic Etching

Metal assisted chemical etching (MACE) is a simple process to achieve anisotropic etching of silicon by utilizing some noble metals as catalysts. A thin layer of noble metal (Au, Pt, or Au/Pd alloy) sputtered on the surface of a Si substrate catalyzed the etching of Si in a mixed solution containing HF and H_2O_2. The Si beneath the noble metal is etched much faster than the Si without noble metal coverage. As a result, the noble metal sinks into the Si substrate, generating pores in the Si substrate. The detailed geometries of the resulting Si structures depend mostly on the initial morphology of the noble metal coverage [12]. The process is illustrated in

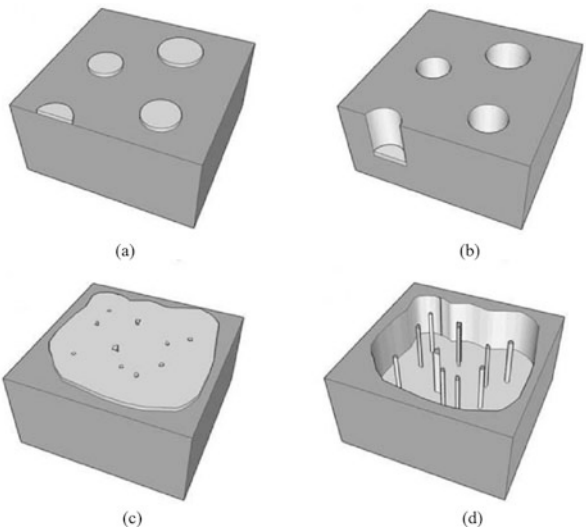


Fig. 7.9 Schematic of metal assisted chemical etching process: (**a**) initial metal discs; (**b**) etched pores with the metal discs at the bottom; (**c**) initial metal film with holes; (**d**) etched pit with standing silicon pillars [13]

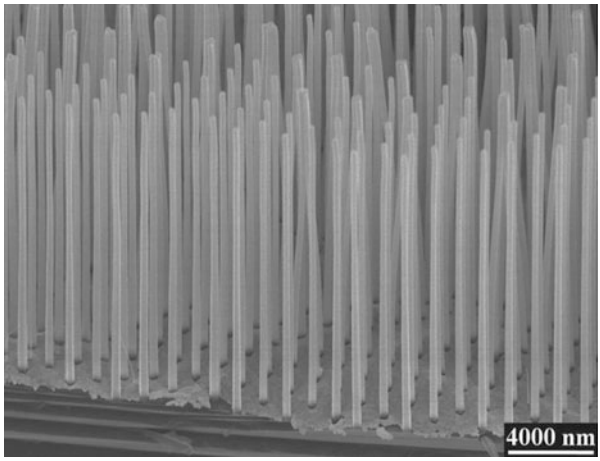


Fig. 7.10 Patterned silicon pillars by MACE [14]

Fig. 7.9a, b where initial metal discs on the Si surface sunk into the Si to form pores. If the metal covers the silicon surface with holes, the end result of MACE is pillars standing in the etched pit, as illustrated in Fig. 7.9c, d [13].

It is generally considered that the noble metals act as catalysts to dissociate the H_2O_2 into oxidants which oxidize the silicon and then the silicon oxide is dissolved in HF acid. Once the silicon oxide is removed, the fresh surface of silicon undergoes new oxidation and etching. The cycle repeats, and the silicon underneath the noble metal is etched away to form deep pores. In this case, the metal film acts as a mask with negative tone, i.e., silicon is etched where there is a metal pattern. A metal pattern can be formed in many ways. For example, Fig. 7.10 shows silicon pillars

made by MACE, which used self-assembled polystyrene (PS) spheres as mask to pattern silver (Ag) film into dots array and used the Ag dots array as masks to etch silicon [14]. Nanoscale silicon structures can be fabricated by MACE with nanoscale noble metal patterns. One of reported work showed vertical silicon nanowires with individual nanowires at 20 nm diameter and height of 4400 nm [15]. Another work demonstrated 25 nm lines with 25 nm spacing were etched in silicon to a depth of 1.26 μm, which has an aspect ratio of ~50:1 [16]. In addition to the well-known noble metal catalysts, such as Ag, Au, and Pt, carbon-based catalysts, such as graphene, graphene oxide, and carbon nanotubes, have been successfully applied as catalysts for MACE, making the process CMOS-compatible. The MACE has also been applied to semiconductors other than Si to fabricate various nanostructures, such as Ge, GaAs, AlGaAs, GaN, GaP, InP, SiC, and Ga_2O_3 [17].

However, as a wet etching technique, MACE suffers from stiction and agglomeration of structures during drying due to van der Waals and capillary forces between adjacent surfaces of dense etched structures. Post etching drying steps, such as CO_2 based critical point drying, are usually required. For this reason, gas phase MACE has been developed. The gas phase of etchant can be created by a vapor of water-diluted hydrofluoric acid and a continuous airflow, which works both as an oxidizer and as a gas carrier for reactive species. The etching of very deep (up to 107 μm) nanostructures (vertically aligned nanowires with diameter in the range of 10~100 nm) was achieved in the Si <100> direction with very high aspect ratio (up to 10,000:1) and high etching rate (up to 25 μm/h) [18].

7.3 Reactive Ion Etching (RIE)

Reactive ion etching (RIE) is the mostly used pattern transfer technique in microfabrication and nanofabrication. The most distinctive feature of RIE is that the etching takes place in plasma [19]. There are no wet chemicals involved. Therefore, the other name for RIE is dry etching, though dry etching includes much broad range of gas phase etching techniques and RIE is only one of them [20].

7.3.1 Principle of RIE

RIE involves both physical and chemical processes. The physical process includes ionization of reactive gas molecules and ion sputtering of sample surface, both of which take place in a plasma. Plasma can be loosely defined as partially ionized gases. In a confined space such as a vacuum chamber, the ionized particles collectively set up a long range electrostatic field which has important influence on the behaviors of these charged particles. The electric field is created because electrons in the plasma tend to move much faster than ions. The fast moving electrons hit the chamber wall before ions do, causing the wall to charge up negatively. The

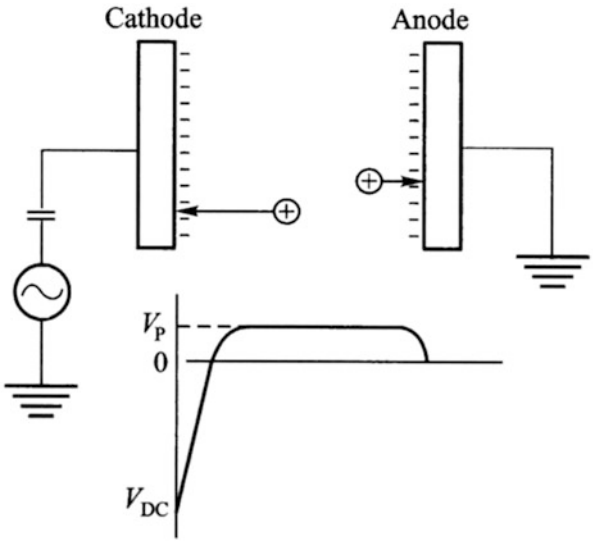


Fig. 7.11 Planar electrode plasma system and potential distribution

negatively charged wall surface pushes other electrons away from the wall, creating a layer of positively charged region where an electrical field is created to accelerate the ions toward the wall surface [21].

In a practical RIE system with planar electrode configuration, as schematically shown in Fig. 7.11, one of the electrodes is the cathode which is connected to radio frequency (RF) power source by capacitance coupling. The other electrode is the anode which is grounded, including the reaction chamber walls. The RF power source provides a high frequency electrical field between the anode and cathode, which accelerates and resonates a small number of free electrons. These electrons acquiring enough kinetic energy in the RF field will collide with gas molecules and ionize the gas molecules. More free electrons become available and more gas molecules are ionized, leading to an avalanche of free electrons and gaseous ions. The gas filled space between cathode and anode changes from initially an insulator to a conductor. Electrical current flows through the space between the two electrodes. The electrical potential distribution between cathode and anode is shown in Fig. 7.11. A large part of the region is equipotential region, which is also called glow region where electrons and ions recombine to form molecules. The gas molecules returning from excitation state to normal state through recombination process release their energy in the form of photons. A glowing light is seen in the plasma chamber. Local electrical fields near the cathode and anode are formed because of the fast moving electrons and charging of the electrode surface. The region near the cathode which is not grounded establishes higher field than that of the anode, which accelerates ions toward the cathode. A sample placed on the cathode electrode will experience bombardment of high energy ions.

Although there is an electrical field which accelerates ions toward the sample surface, the physical sputtering is not a dominant phenomenon in RIE. The dominant

Table 7.1 Materials and corresponding RIE gases

Materials to be etched	Chemical gases (multi-choices)
Single crystal silicon	CF_3Br, HBr/NF_3, SF_6/O_2
Polysilicon	$SiCl_4/Cl_2$, BCl_3/Cl_2, $HBr/Cl_2/O_2$, HBr/O_2, Br_2/SF_6
Al	$SiCl_4/Cl_2$, BCl_3/Cl_2, HBr/Cl_2
Al-Si-Cu, Al-Cu	$BCl_3/Cl_2 + N_2$
W	SF_6, NF_3/Cl_2
TiW	SF_6
WSi_2, $TiSi_2$, $CoSi_2$	CCl_2F_2/NF_3, CF_4/Cl_2
SiO_2	CCl_2F_2, CHF_3/CF_4, CHF_3/O_2, CH_3CHF_2
Si_3N_4	CF_4/O_2, CF_4/H_2, CHF_3, CH_3CHF_2
GaAs	$SiCl_4/SF_6$, $SiCl_4/NF_3$, $SiCl_4/CF_4$
InP	CH_4/H_2
Photoresists	O_2

etching process is caused by chemical reaction. Most RIE processes use halogen-containing gases, such as fluorine (F), chlorine (Cl), bromine (Br), and iodine (I), to generate the plasma, because they are the most chemically reactive and erosive gases to various inorganic materials. Fluorine is widely used for etching silicon and silicon compounds. Chlorine is mainly used for etching Ga-, Al-, and As-containing III–V compounds. For InP and other In-containing compounds such as InGaAs and GaInAsP, the volatility of etched product in chlorine plasma at room temperature is low. Either the substrate has to be heated to more than 150 °C or bromine- or iodine-containing gases are used (but such gases tend to attack the vacuum system aggressively) [22]. There are, however, cases that non-halogen-containing gases (hydrogen or hydrogen/methane mixture) were used to etch compound semiconductors such as GaAs and InP [23].

The halogen compounds that can be used in RIE and the materials which can be etched are listed in Table 7.1 [24]. One can notice that for etching of some of the materials, there are several choices of etching gases. The gases listed in Table 7.1 only serves as references. There can be many more choices. Each gas has its own advantages and drawbacks. The choice of a particular etching gas is very much dependent on the requirements of a specific application. For example, fluorine compounds and chlorine gas or compounds can both be used for etching silicon. F atoms react with Si spontaneously, and they induce isotropic etching and cause undercutting below the mask. Cl_2 etching is anisotropic since it is an ion assisted process rather than a spontaneous etching process. However, Cl_2 etching has much lower etch rate and poorer selectivity than fluorine based etching [25].

The compound forms of halogen gas, though chemically erosive, are rarely able to etch a material or to etch at very low rate. They only become chemically reactive and erosive when these halogen atoms are dissociated from the compounds or lose their electrons to become free radicals. Here the energetic ions play an important role to produce the radicals. An interesting experiment demonstrated that when silicon surface was exposed to XeF_2 gas or Ar^+ beam alone, the silicon was removed at less

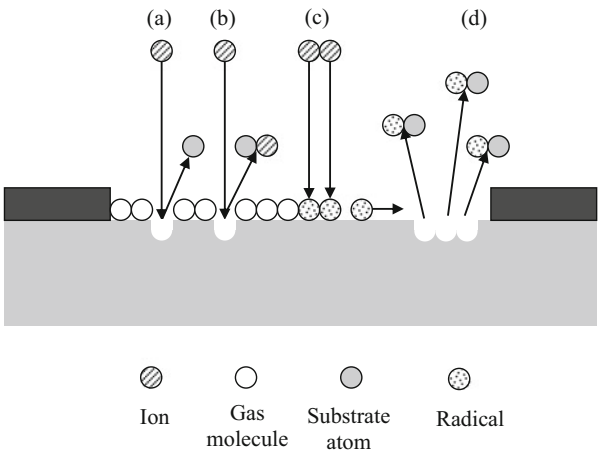


Fig. 7.12 Schematic of RIE process: (**a**) ion sputtering; (**b**) reactive etching; (**c**) radical formation; (**d**) radical etching

than 1 nm/min. However, if the XeF_2 gas and Ar^+ beam were both present, the etch rate jumped to 6 nm/min [22]. The RIE process is the action of four simultaneous processes, as schematically shown in Fig. 7.12 [22]:

(a) The ions sputter the substrate surface, which helps initially to remove any hydrocarbon contaminants and native oxide layer to create a fresh material surface, promoting absorption of gas molecules onto the material surface. Ion sputtering can also remove material atoms as purely physical sputtering etching.
(b) The ions, if they are reactive species, can directly react with material atoms to form volatile compounds which escape from the substrate surface.
(c) The ions can dissociate the absorbed gas molecules to produce radicals on the substrate surface.
(d) The radicals can move around at the substrate surface and react with the material surface atoms to form volatile compounds leaving the surface. All the volatile compounds are pumped away, so that no redeposition occurs during the etching process.

It is clear that there are three factors that lead to the RIE: the energized ions, the reactive gas, and the formation of volatile compounds. First, if there are only energized ions without reactive gas, for example, of etching in pure Ar plasma, then the etching is pure physical sputtering. Ion milling is a typical sputter etching technique conducted in argon plasma. Second, if there are only reactive gas molecules without energized ions, the radicals are only generated in plasma not on the surface. The etching is a pure chemical reaction. For example, if the sample is placed on the anode shown in Fig. 7.11, there is little electrical field to accelerate the ions. Or there are no electrodes at all, such as in a barrel etcher, the ion bombardment of surface is negligible. The etching is isotropic and similar to aforementioned wet chemical etching. The third factor is the volatility of etched products. If there are no

volatile compounds formed during etching, the removed material by etching will redeposit somewhere else in the vicinity of etched pattern.

The volatility of halogen compounds can be characterized by their boiling temperature. The boiling temperatures of halogen compounds for some most commonly encountered elements in micro- and nanostructure fabrication are listed in Table 7.2. [26]. The data in Table 7.2 can serve as a guideline for selecting appropriate halogen gas when etching a particular material. The general trend is that when going from fluorine to chlorine and bromine chemistry, etch rates decrease and etching needs more ion assistance. An example has already been described previously about etching silicon with fluorine and chlorine gases. Fluorine plasma etches silicon faster but is less anisotropic. Chlorine plasma etches silicon slower but is more anisotropic. This behavior, which is in favor of an easier profile control, is in part due to the reduction of the etch product volatility. For compound materials such as GaAs, there is generally not one etch product for the compound material, but one etch product per constituent, each having its own volatility. This complicates the choice of the chemistry and has strong effects on the etching mechanisms and on the chemical composition of the etched surface. To increase the volatility of those etched products with high boiling temperatures, the sample has to be heated to high temperature. In some cases, adding hydrogen gas into plasma works better because some hydrides have much higher volatility than the halogen compounds, as shown in Table 7.2.

7.3.2 Process Control in RIE

RIE is a complex process. The above description is only a qualitative and simplistic description of what is happening inside the plasma etching chamber. So far there are no theories that can satisfactorily describe the complete RIE process, though there have been various attempts to establish theoretical models of plasma etching processes [21]. Carrying out pattern transfer by RIE is highly experimental and empirical. There are RIE parameters which can be externally controlled, such as gas flow rate, RF power, chamber pressure, substrate temperature, substrate bias voltage, etc. There are also measurable indicators that indicate the quality of etching, such as the aforementioned etching selectivity and anisotropy, as well as etch rate and etch uniformity. The control of an RIE process is to adjust these parameters experimentally to match them with a desirable outcome from the above indicators. Though theoretically predicting a plasma etching process is difficult, there are some broad guidelines with which a user should be able to anticipate what may happen when certain process parameters are changed. These guidelines are briefly described in the following:

(1) Flow rate of reactive gas

The basic process in RIE is gaseous ions reacting with substrate surface and converting the surface material into volatile products. The reaction rate is directly

Table 7.2 Volatility of etched compounds with different gas chemistries

Elements	Fluorides	Boiling temperature (°C)	Chlorides	Boiling temperature (°C)	Bromides	Boiling temperature (°C)	Hydrides trimethyls	Boiling temperature (°C)
Al	AIF_3	1297 (subl.)	$AlCl_3$	178 (subl.)	$A1Br_3$	263		
As	AsF_3	−63	$AsCl_3$	130.2	$AsBr_3$	221	AsH_3	−55
	AsF_5	−53			$AsBr_5$			
C	CF_4	−128	CCl_4	77	CBr_4	189	CH_4	−164
Cr	CrF_2	>1300	CrO_2Cl_2	117	$CrBr_2$	842		
Cu	CuF	1100 (subl.)	CuCl	1490	CuBr	1345		
	CuF_2	950	$CuCl_2$	993			CuH	55–60
Ga	GaF_3	1000	$GaCl_3$	201.3	$GaBr_3$	278.8	$Ga(CH_3)_3$	134
Ge	GeF_4	−37 (subl.)	$GeCl_4$	84	$GeBr_4$	186.5	GeH_4	−88.5
In	InF_3	>1200	$InCl_3$	300 (subl.)			$In(CH_3)_3$	55.7
Mo	MoF_5	213.6	$MoCl_5$	268				
	MoF_6	35	$MoOCl_3$	100 (subl.)				
	MoO_2F_2	270 (subl.)						
	$MoOF_4$	180						
P	PF_3	−101.5	PCl_3	75	PBr_3	172.9	PH_3	−87.7
	PF_5	−75	PCl_5	162 (subl.)	PBr_5	106		
Si	SiF4	−86	$SiCl_4$	57.6	$SiBr_4$	154	SiH_4	−111.8
Ta	TaF_5	229.5	$TaCl_5$	242	$TaBr_5$	348.8		
Ti	TiF_4	284 (subl.)	$TiCl_4$	136.4	$TiBr_4$	230		
W	WF_6	17.5	WCl_6	346.7				
	WOF_4	187.5	WCl_5	275.6	WBr_5	333		
			$WOCl_4$	227.5	$WOBr_4$	327		

Reproduced from [26] with permission

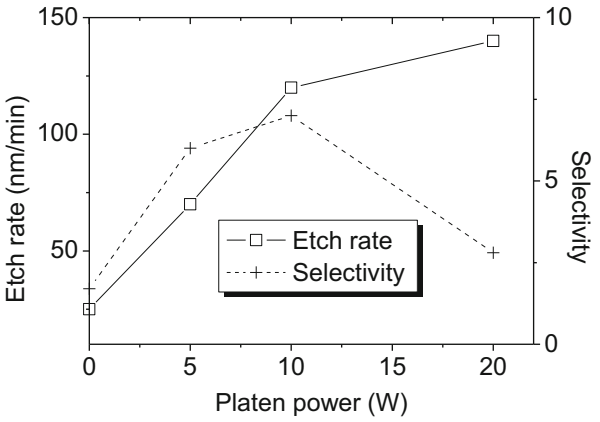


Fig. 7.13 Influence of platen power on Si etch rate and etch selectivity

linked to the supply rate of reactive gases. However, fast flow rate of reactive gases at a constant pressure can lead to short dwelling time of gas molecules in reaction chamber, because the constant pressure is maintained by high pumping speed. If plasma discharge power is not increased accordingly, the number of gaseous ions generated in reaction chamber will decrease, resulting in reduction of etching rate. On the other hand, if the gas flow rate is too low, the consumed reactive gases cannot be replenished sufficiently, which again can lead to drop of etching rate.

(2) RF power

Increasing RF power will increase electron energy, hence increasing ionization probability. Therefore, etching rate generally increases with the increase of RF power, particularly when gas flow rate has increased or chamber pressure has increased. As explained earlier, high energy electrons not only generate more ions but also build up a high negative field near the cathode surface, called self-bias field, which accelerates ions toward cathode. The potential difference at cathode region is 200–500 volts. Accelerated ions bombarding cathode (platen) where the sample is situated can also increase the etching rate. In addition, the anisotropy of etching can be improved because ion sputtering is highly directional and almost has no lateral etching. However, ion sputtering is a physical process, irrespective of mask or substrate. Too much of ion sputtering will speed up the consumption of masking layer and reduce etching selectivity. Figure 7.13 shows the Si etch rate and selectivity as functions of platen power which represents the energy of bombarding ions in an induction coupling plasma etching system [23]. Although high platen power is favored for high etch rate, the etch selectivity suffers at the high platen power.

(3) Chamber pressure

RIE is normally taking place at low pressure condition. The chamber vacuum is usually at $10^{-3} \sim 10^{-1}$ Torr. At low pressure, the density of gas molecule is low in the chamber. Electrons have longer mean free path and can acquire more acceleration energy between two collisions with gas molecules, which can increase

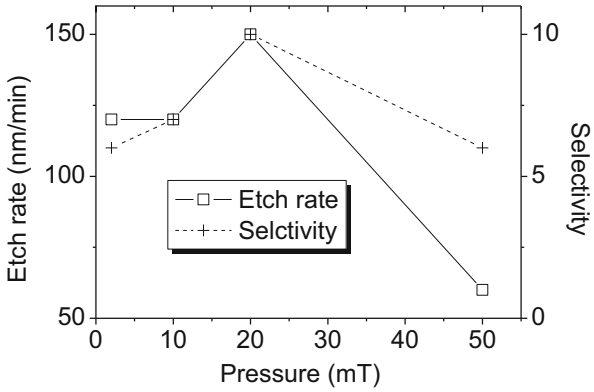


Fig. 7.14 Influence of chamber pressure on Si etch rate and selectivity

ionization probability. Another benefit of low chamber pressure is less lateral etching and improvement of anisotropy. This is due to fewer collisions between ions and gas molecules, resulting in more directional movement of ions toward substrate. Low chamber pressure also means high pumping speed, which helps to quickly move away the volatile products from etched surface. The reaction rate is improved, hence the etching rate. In general, at low chamber pressure, RIE is dominated by ion sputtering etching, because chemically active gaseous molecules are less and ion energy is higher. At high chamber pressure, the etching is dominated by chemical reaction, because the mean free path of ions is reduced and the number of chemically active ions is large. At higher pressure, the etching is more isotropic. However, if the pressure is too low, the plasma cannot start or sustain. There exists an optimum pressure at which both etch rate and selectivity are maximum, as an example shown in Fig. 7.14, where both Si etch rate and selectivity are decreased when chamber pressure is too low or too high [23].

(4) Substrate temperature

In general, high temperature can promote chemical reaction and also help to remove reacted products from substrate surface. As described earlier, etching of indium-containing III–V semiconductor compounds with chlorine plasma requires high substrate temperature, because of the low volatility of indium-chlorine compounds. In RIE, ion bombardment of substrate will naturally cause the rise of substrate temperature. The substrate surface temperature can reach 100–200 °C, depending on the duration of etching. However, high substrate temperature can soften masking material if it is photoresist, leading to pattern distortion and higher photoresist etch rate. Using other inorganic materials instead of photoresist as mask will avoid the problem. Another effect of high substrate temperature is the increase of lateral etching, because the mobility of radicals is enhanced at high surface temperature and hence the chemical etching which is by nature an isotropic etching. For long duration of etching, the sample holder has to be cooled to reduce the substrate temperature.

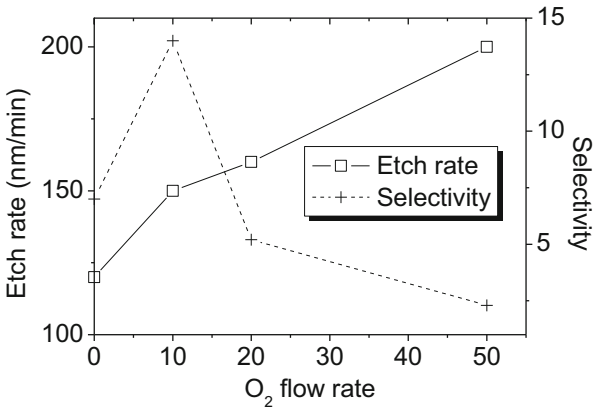


Fig. 7.15 Influence of oxygen flow rate on Si etch rate and selectivity

(5) Gas additives

Although halogen compounds are the main gases for etching inorganic and metallic materials, additives are often used to enhance some aspect of the chemistry. For example, the addition of 10% O_2 to CF_4 increases the etch rate by ten times. In the plasma, CF_4 is broken down into radicals CF_3 and CF_2. In the absence of oxygen, these radicals can recombine (both in the gas phase and on surfaces) into stable molecules, e.g.,

$$CF_3 + CF_3 \rightarrow C_2F_6 \text{ and } CF_3 + F \rightarrow CF_4$$

With oxygen present, it can react with the radicals to form CO, CO_2, and COF, thus releasing fluorine radicals:

$$CF_2 + O_2 \rightarrow COF + F, COF + O \rightarrow CO_2 + F$$

Figure 7.15 shows the Si etch rate and selectivity as the functions of flow rate of O_2 additive [23]. Addition of oxygen has clearly enhanced the silicon etch rate. However, the selectivity is substantially reduced because oxygen is an effective etching gas to etch polymer resists which is the masking material in this case. Nevertheless, addition of oxygen also brings the benefit of eliminating micromasking effect and helps to produce smooth etched surface.

Apart from oxygen gas, argon gas is often added into plasma. While Ar + can participate in physical sputter etching, they can also help to stabilize the plasma. For example, SF_6 is an electronegative gas. It absorbs electrons upon ionization, which may reduce the total number of electrons in the plasma. The addition of argon can add electrons to the plasma when they are ionized and so helps to produce a stable plasma [22]. However, adding too much argon can dilute the concentration of reactive gas and also cause excessive physical sputtering, which leads to reduction of both etch rate and selectivity.

On the one hand, RIE is a complex process and is influenced by many parameters. On the other hand, there are more degrees of freedom to control the process conditions and to optimize. It is possible to fine-tune process parameters to reach a compromise between etch rate and selectivity and to achieve a particular etch profile. Figure 7.16a–c are three different profiles of etched silicon tips. Although the basic etching chemistry is based on SF_6 plasma, different shapes of etch profile can be fine controlled by adding CHF_3, O_2, and Ar gases [27]. The sharp end of these silicon tips can be controlled to less than 50 nm in diameter.

The control of anisotropy in RIE is tricky because etching by halogen radicals in plasma is nondirectional. It causes the etching to advance both in depth and laterally, resulting in the pyramid shape of etched profile as shown in Fig. 7.16. Only the ion bombardment of substrate is directional. However, halogen radical etching is very sensitive to polymers. A thin layer of polymer or carbon compounds can stop the etching. A rough guide to control the anisotropy of RIE is to find a way to create a passivation film on the etched sidewall to stop halogen radicals etching laterally, while any passivation on the bottom of etched structure can be removed by directional ion bombardment with the progressing of etching process. It is not difficult to find carbon compounds in RIE because most of the etching processes employ polymer resists as the etching mask layer. The carbon compounds in the plasma can deposit onto the sidewall of etched structures, forming the passivation layer. It is also possible to form the passivation layer by introducing carbon compound gas, such as CHF_3, into the plasma. Sub-50 nm silicon nanowires have been produced with aspect ratio up to 50:1 using mixture of SF_6 + CHF_3 etching gases [28]. With the assistance of CHF_3, sub-100 nm silicon structures can be etched in RIE even without the use of a polymer masking layer. Figure 7.17 shows etched silicon nanostructure of 50 nm in width and 350–400 nm in height with near vertical sidewall profile in which 10 nm chrome film was used as masking layer. The etching was done with CHF_3 and SF_6 gas mixture at RF power of 250 W and chamber pressure of 30mTorr [29].

7.3.3 RIE by Inductively Coupled Plasma (ICP)

It has been demonstrated that conventional RIE is perfectly capable of etching sub-100 nm structures. The only limitation is its relatively low etch rate, which is normally less than 200 nm/min for silicon [30]. The etching rate in RIE directly depends on plasma density. For a conventional RIE system as shown in Fig. 7.11, plasma density increases with increasing RF power. However, increasing RF power will also increase the self-biasing voltage on the cathode where the etching sample is situated. The consequence is the increase of ion bombardment energy, hence deterioration of etching selectivity. This becomes a particularly serious problem in sub-100 nm RIE, because the masking layer is always thin to enable photon or e-beam patterning of polymer resist at sub-100 nm feature dimension. It is also a

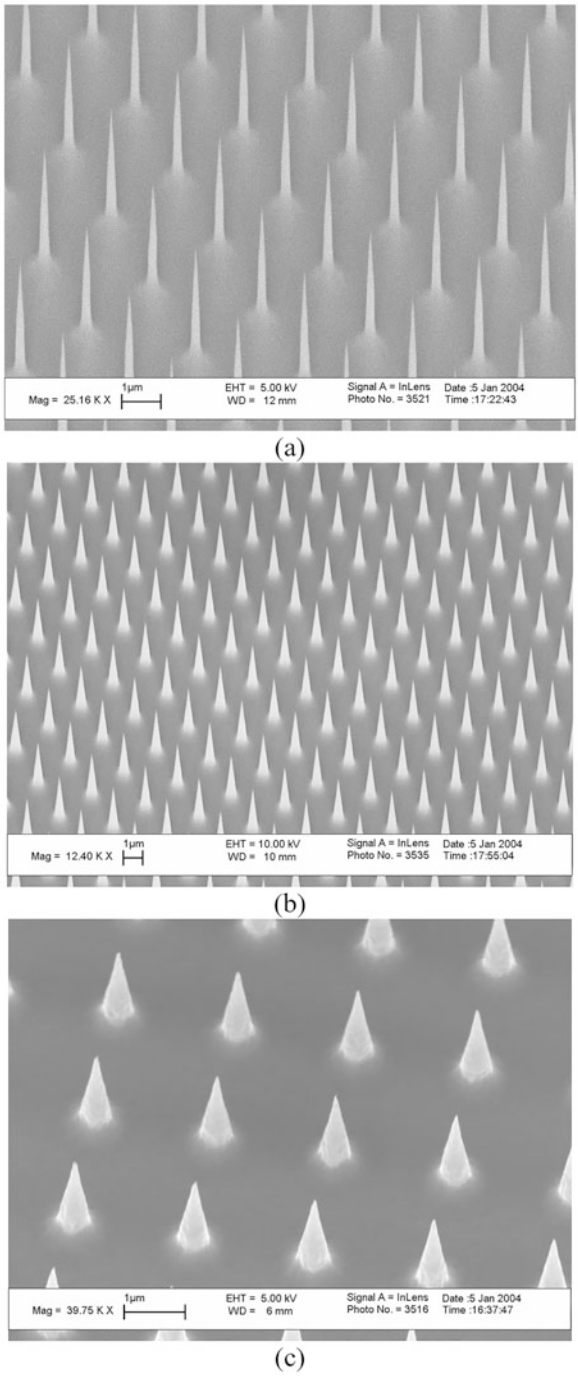


Fig. 7.16 Controlled etching to achieve different shapes of silicon tips from very sharp tips (**a**) to medium sharp tips (**b**) and to blunt tips (**c**)

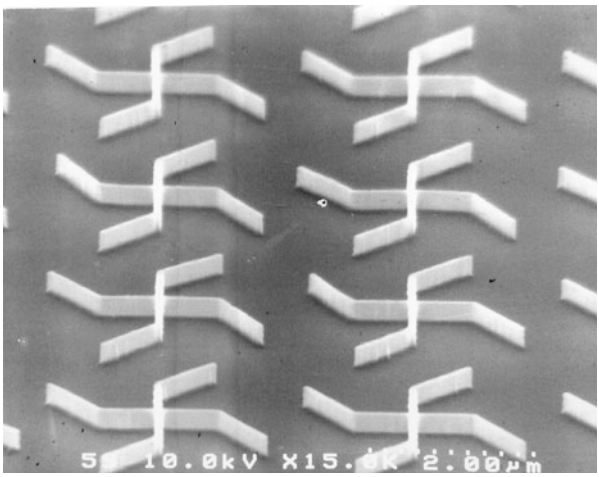


Fig. 7.17 Sub-100 nm etched silicon structure using RIE with CHF_3 and SF_6 gas mixture and chrome film as etching mask

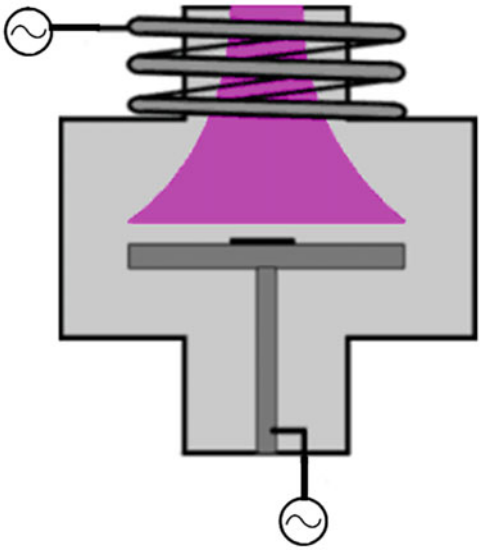


Fig. 7.18 Schematic of inductive coupling plasma source

serious issue for deep structure etching as longer etching at high etching rate consumes masking layer faster.

Inductively coupled plasma (ICP) system has cleverly solved the problem. In an ICP source, the plasma generation is separated from etching chamber, as shown in Fig. 7.18. RF power is coupled into plasma chamber by an induction coil from outside. The sample stage is connected to a second RF power source as an auxiliary RF source to enhance the production of plasma. Electromagnetic field generated by inductive coupling coil can sustain electron cycling movement in plasma for a long period, which has greatly increased ionization probability while keeping the pressure

low in the etching chamber. As the sample stage has an independent input of RF power, the self-biasing voltage can be independently controlled. Therefore, an ICP system can produce very high plasma density ($>5\times10^{11}/cm^3$) compared to conventional RIE (10^8–$10^{10}/cm^3$) and maintain low ion bombardment energy. The conflict between high plasma density and high etch selectivity encountered in conventional RIE systems has been resolved. The current best ICP systems can produce about 20 μm/min silicon etching rate which depends on the area size to be etched on silicon wafer. The selectivity between silicon and photoresists can be as high as 100:1. Selectivity between silicon and oxide is over 200:1. With such high plasma density and high etching selectivity, high etching rate and deep reactive ion etching (DRIE) become possible.

Because of its high density plasma, ICP is not only widely used for deep silicon etching but also an efficient etching tool for other materials, such as quartz [31], III–V semiconductors [32], silicon carbide [33], and even tungsten which can be etched to the depth of more than 300 μm at an etch rate of 2.73 μm/min using patterned aluminum as the etch mask [34]. Commercial ICP etchers are readily available from a number of vendors worldwide [35–38].

7.4 Deep Reactive Ion Etching (DRIE)

To fabricate silicon MEMS (microelectromechanical systems), the bulk of silicon needs to be removed. Bulk silicon micromachining is used to be done with alkaline wet etching as described in Sect. 7.2.2. With ICP technology, bulk silicon micromachining as well as a range of other materials becomes possible by RIE due to its high etch rate. Most MEMS devices not only require to remove large quantity of silicon but also require high aspect ratio structures to enhance their electromechanical performance. However, conventional RIE is inherently not highly anisotropic. There are always lateral etch when the etch goes deeper. Deep reactive ion etching (DRIE) requires high anisotropy to achieve high aspect ratio etched structures.

In addition to MEMS applications, semiconductor industry also requires high aspect ratio silicon etching to fabricate through-silicon-via (TSV). TSVs are deep holes etched into silicon and filled with conductive metals to serve as electrical interconnects. The primary benefit of TSVs is the reduced interconnect length. The short vertical connections through thinned silicon die enable reduced latency, lower capacitance, lower inductance, and permit higher speed communications, higher numbers of interconnections, and lower power level of communication links between circuits. With the help of TSVs (Sect. 11.2 of Chap. 11), thinned die and stacked die are possible, which permit miniaturization of integrated multi-chip systems and have enabled today's mobile communications and system-on-chip (SoC) ICs [39]. For 3D flash memory (NAND) chips, DRIE has become an essential fabrication technology [40].

7.4.1 Bosch Process

To ensure high anisotropy of DRIE, a gas switching process called Bosch process was invented in 1992 [41], which deliberately passivates the sidewall to stop lateral etching. The typical Bosch process conditions are listed in Table 7.3, and its steps are schematically illustrated in Fig. 7.19. It is a repeated etching-passivation process by introducing alternating gases. The etching gas is SF_6 and passivation gas is C_4F_8. C_4F_8 in plasma can form a fluorine carbon compound which once deposited on silicon surface can prevent F^+ from further reacting with silicon. Etching and passivation are switched over every 5–15 seconds a cycle. Although pure SF_6 gas reacting with silicon results in isotropic etching, the etched surface is soon passivated by C_4F_8 after a few seconds of C_4F_8 reaction. In next round of SF_6 gas being introduced, ion bombardment in depth direction removes the passivation film to

Table 7.3 Typical Bosch process conditions

Process parameters	Passivation	Etching
C_4F_8	85 sccm	0 sccm
SF_6	0 sccm	130 sccm
RF power at stage	0 watt	12 watt
RF power from coil	600 watt	600 watt
Cycle time	7.0 s	9.0 s
Delay time	0.5 s	0.5 s
Etch rate		1.5–3 μm/min

sccm standard cubic centimeter per minute

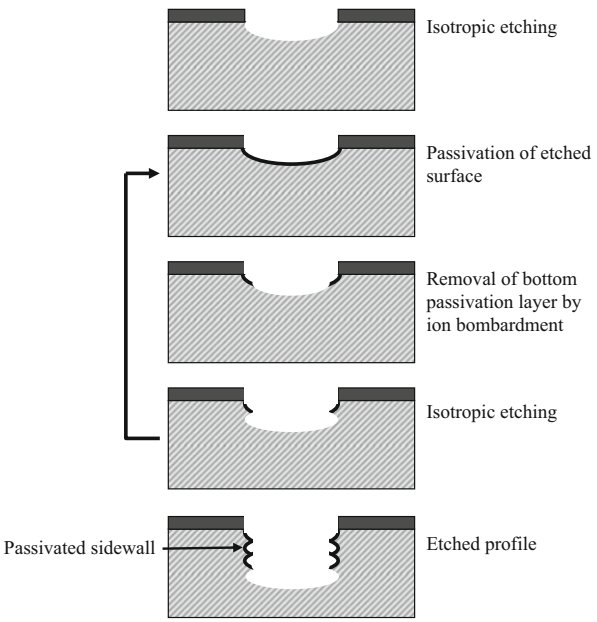


Fig. 7.19 Schematic of Bosch process

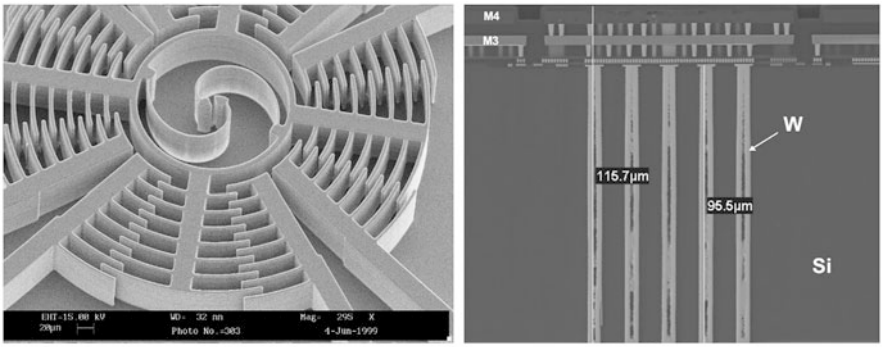


Fig. 7.20 Examples of Bosch process: (**a**) a silicon MEMS gyroscope; (**b**) cross-sectional image of TSVs filled with tungsten in a silicon wafer (reprinted from [39] with permission)

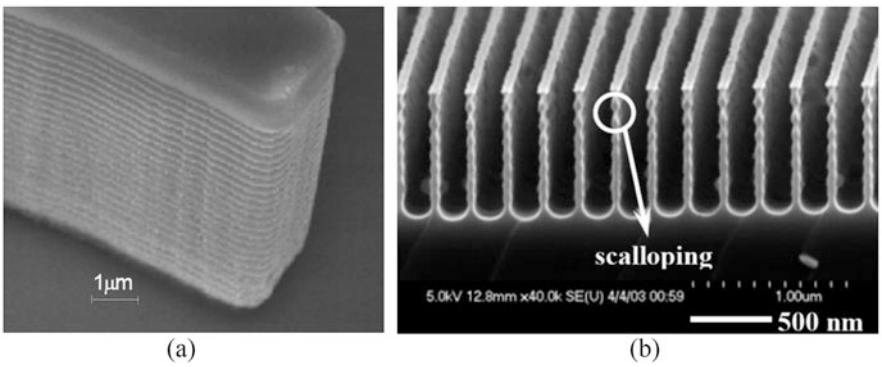


Fig. 7.21 (**a**) Scalloping effect created in "Bosch" process; (**b**) scalloping effect in etched sub-100 nm silicon structures (reprinted from [43] with permission)

allow SF_6 reacting further with the fresh silicon surface, whereas ions cannot sputter the sidewalls. The passivation film remains intact to protect sidewalls from being etched. Through the periodical etching-passivation-etching cycle, silicon etching only proceeds in depth direction and eventually forms vertical sidewalls. In this way, highly anisotropic etching can be achieved, as the examples of etched MEMS and TSVs shown in Fig. 7.20.

Though it is capable of etching very deep structures with near vertical sidewall profiles, the Bosch process has a fundamental problem: the scalloping effect, e.g., ripples are formed over the sidewall surface due to the switching between etching and passivation, as the example shown in Fig. 7.21a. The surface roughness due to the scalloping is in the order of more than 100 nm [42]. The surface roughness caused by the scalloping effect may be tolerable if the feature dimension is tens or hundreds of micrometers, which is the case in MEMS devices. However, it cannot be

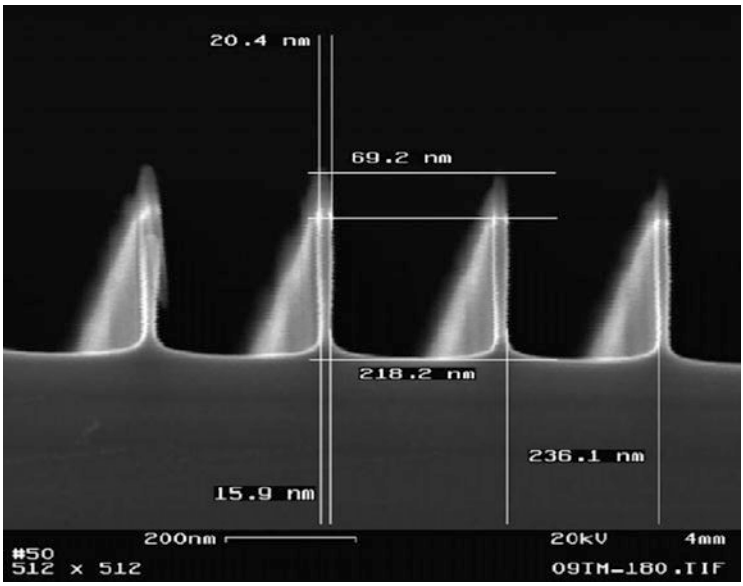


Fig. 7.22 High aspect ratio silicon nanostructures etched by non-switching DRIE process (16 nm line width) (reprinted from [48] with permission)

tolerated if the feature dimension is in sub-100 nm range. An example is shown in Fig. 7.21b where the passivation time is 5 seconds and etching time is 6 seconds. There are eight cycles of gas switching, which can almost be counted from the number of scallops on the sidewall [43].

Many investigations have been carried out to reduce the scalloping effect. For large structure dimension, the scalloping may be smoothed by post etch polishing in wet chemicals such as KOH + IPA [44, 45]. For sub-100 nm feature dimension, the wet chemical polishing may distort the structure. Other techniques have been developed, including prolonging the passivation time [46], using a Teflon coated sample bed [47], or simply mixing the SF_6 etch gas with the C_4F_8 passivation gas without switching. Figure 7.22 shows etched 16 nm silicon line structures using the non-switching $SF_6 + C_4F_8$ DRIE process [48]. By optimizing the etching parameters and process conditions, a simple mixing of SF_6 and C_4F_8 gases has also achieved sub-10 nm vertical silicon nanowires with aspect ratio of ~50:1 and smooth sidewalls [49]. Smooth sidewalls can be achieved even with standard ICP etchers by modifying both SF_6 pulse time and lowering the substrate holder temperature. The sidewall scallop depth achieved by a modified Bosch process was reduced from 271 nm to 42 nm via the generic process optimization [50].

Nonetheless, the typical surface roughness by Bosch process is in excess of 200 nm. However, when etching narrow trenches (in the range below 5 μm wide), this roughness drastically reduces to levels below 25 nm, and it eventually vanishes with increasing aspect ratios during the process. Very high aspect ratios up to 160:1 on trenches of 250 nm width have been achieved using the Bosch process, which

indicates that surface roughness induced by the cyclic nature of the Bosch process is significantly attenuated when increasing the aspect ratio of deep trenches, which is especially true for the submicron trenches [51].

7.4.2 Cryogenic Process

Another sidewall passivation technique is low temperature etching of silicon, also called cryogenic process. During etching of silicon, SF_6 gas reacting with silicon forms volatile compound SiF_4. At room temperature, the compound is quickly removed from reaction chamber by vacuum pumping system. However, at low temperature (<-100 °C), the compound can form a 10–20 nm protective film over silicon surface, such that the reaction of fluorine free radicals with silicon is greatly weakened because of this protective film. Also weakened is the reaction with photoresist or oxide. Both the sidewall protection and high selectivity of mask to silicon are realized by the cryogenic process, resulting in excellent anisotropic etching of silicon [52]. Introducing oxygen into the low temperature chamber can also help to form a protection layer on the sidewall, which prevents the sidewall from lateral etching.

The use of photoresist masking layers is difficult with cryogenic etching since the resist is a polymer and can crack with the low temperatures. For this reason, oxide mask is preferred. The mask selectivity to silicon is approximately 150 to 1 for a silicon dioxide and 70 to 1 for photoresist [53]. The advantage of cryogenic etching for deep structure is that there is no scalloping of the sidewalls of the etched features. The optimum low temperature is -110°C~-140 °C, at which no lateral etching of silicon was found. Below -140 °C, the entire silicon etching stops because the etching gas SF_6 starts to freeze on the cold substrate surface [54]. The main disadvantage of cryogenic etching is the low process temperature. Since the substrate is at such a low temperature, it attracts particulates and contaminates in the process chamber that can lodge onto the surface, thereby creating micromasking effects as the defects in the etched structures.

Table 7.4 compares the performances of three non-switching DRIE process for etching silicon nanostructures: HBr process, room temperature SF_6 + C_4F_8 process (RT F-base), and cryogenic process [48]. Each process has different etch rate and selectivity, but all are capable of producing very smooth sidewalls (<10 nm roughness).

7.4.3 Critical Issues in DRIE

When using ICP-based DRIE for pattern transfer, there are a number of critical issues one has to be aware. These issues include loading, microtrenching,

Table 7.4 Comparison of three non-switching DRIE processes

Process	HBr	RT F-base	Cryogenic
Depth	0.05–1 μm	0.05–10 μm	0.2–>100 μm
Feature size	≥5 nm	≥5 nm	≥5 nm
Aspect ratio	≥5:1	≥5:1	≥10:1
Etch rate (nm/min)	≥100	≥200	≥300
Uniformity	<±5%	<±5%	<±5%
Selectivity over oxide	>100:1	>10:1	>30:1
Selectivity over resist	>3:1	>5:1	>15:1
Profile	80–92°	80–92°	80–92°
Sidewall roughness	<10 nm	<10 nm	<10 nm

Reproduced from [48] with permission

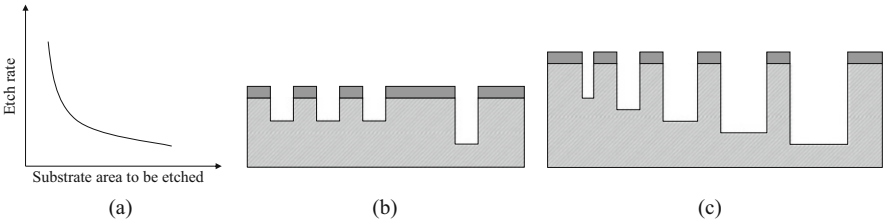


Fig. 7.23 Loading effects in DRIE: (**a**) macroloading; (**b**) microloading; (**c**) aspect ratio dependent etching

micromasking, notching, and ion induced damage, which will directly influence the etch uniformity, control of etch profile, or damage of substrate materials.

7.4.3.1 Loading Effect

Reactive ion etching rate depends on the supply rate of chemically active species. When the consumption of radicals surpasses the supply of them while the gas flow rate and other conditions remain the same, the etching rate will decrease, causing the so-called loading effect. There are three types of loading effects—macroloading, microloading, and aspect ratio dependent etching (ARDE) [55, 56]—as illustrated in Fig. 7.23. They all mean the same thing, i.e., nonuniform etching due to the difference in supply and consumption of reactive radicals at the etching surface.

Macroloading effect is the decrease of overall etch rate due to the increase of total area of materials to be etched away, as shown in Fig. 7.23a, which poses the problem that a well-characterized etching process for one set of the pattern design may not be applicable to another set of design if the overall etching areas are different.

Microloading effect is related to localized depletion of reactive species or accumulation of etching production as a result of local variation of pattern density, as schematically illustrated in Fig. 7.23b. High density pattern areas demand high rate

supply of reactive species as well as high rate removal of etched volatile compounds than low pattern density areas. Such differences in the rates of radicals supply and etched product removal can cause difference in etch depth between high density pattern area and low pattern density area. Generally, isolated patterns are etched deeper than dense patterns.

ARDE is strongly dependent on feature dimensions. Other names for the ARDE are etch lag or aperture effect. As illustrated in Fig. 7.23c, the smaller the feature dimension (or pattern opening to etching), the less is the etch depth under the same etching condition. The cause of ADRE is very complicated. There are as many as eight different mechanisms that may explain the ADRE phenomenon [55], but it is believed to be mostly determined by the mean free paths of the species since this has a significant impact on their transport in narrow trenches. The most direct method for modifying the mean free path is to adjust the process pressures. Since the Bosch DRIE process is composed of two parts wherein the pressures of each part can be individually controlled, the etch lag can be reduced by increasing the process pressure during the passivation step relative to the pressure during the etch step. Additionally, it has been found that decreasing the temperature of the substrate during processing from 40 °C to 0 °C can enable the etch to exhibit nearly lag free behavior over a range of aspect ratios [57].

Loading effect is most significant in DRIE because of prolonged etching and large etching depth. The longer the etching time, the more significant is the loading effect. There can be 100% variation in etching depth [56]. Conventional RIE also has the problem of aperture effect, when both large and small features have to be etched on the same substrate. For example, the etch rate of SiO_2 has shown to drop from 32 nm/min to 12 nm/min when the etched feature size changes from 20 μm to 52 nm [58]. There are no very effective measures that can eliminate the loading effect. Generally, high pumping speed (low chamber pressure), diluting gases, or using pulsed plasma can reduce to some extent the loading effect [59]. These measures help to remove the etched products fast or to reduce local nonuniformity of reactive species. Alternatively, the pattern layout can be modified to reduce or eliminate the nonuniformity in feature density or feature dimension, for example, introducing dummy features purely to balance the feature density or subdividing large features in multistep etching [45]. If the pattern design cannot be easily altered, one has to decide what is the critical feature dimension in a design layout if different feature sizes coexist in the design. Then the RIE condition should be optimized for the critical feature dimension, allowing variation in etch depth for other noncritical features.

7.4.3.2 Microtrenching Effect

Microtrenching is a phenomenon that the etch rate is larger near the corners of a trench compared to the center of the trench. Typical microtrenching in etched features is shown in Fig. 7.24a where deep grooves appear adjacent to the sidewalls of etched structure. Microtrenching is believed to be produced by the impact of high

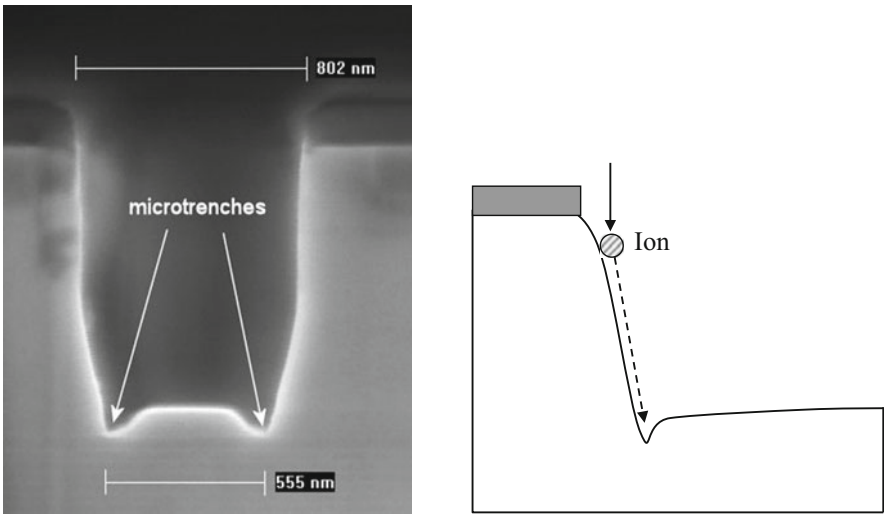


Fig. 7.24 (**a**) Microtrenching effect in RIE; (**b**) mechanism of microtrenching

energy particles (mostly ions) at grazing angles (>80°) on the sidewalls and then reflected toward the bottom of the etched surface, as illustrated in Fig. 7.24b. These ions retain a large fraction of their energy and directionality. It is like a "focusing" of the high energy particles at the base of the sidewalls of the feature, resulting in higher etch rates at those locations [60]. Such a mechanism has been proven by Monte Carlo simulation of ions sliding at sloped sidewalls [61]. Both sidewall slope angle and incident angle of ions can significantly influence the microtrenching formation. Generally, the deeper the etch depth, the more significant is the microtrenching because a larger sidewall surface becomes available for reflecting ions. Negative charging of sidewalls can also contribute to microtrenching, because the incident ions can be deflected by the negative field toward the sidewall [62]. By deliberately introducing flux of electrons into the etched trenches, the local electrical field can be neutralized and the microtrenching effect can be reduced [63].

7.4.3.3 Notching Effect

Notching is the phenomenon of lateral etching at the foot of etched structure, which occurs when etching nonuniform size of features in silicon-on-insulator (SOI) substrates. Typical notching effect is shown in Fig. 7.25a [64]. SOI wafers are special silicon substrates in which a thin SiO_2 layer is sandwiched between a thick silicon wafer (handle layer) and a thin silicon wafer (device layer) [65]. Theoretically, the SiO_2 insulator underneath the device layer provides an excellent etch stop layer which ensures all features, no matter big or small, are all etched to the same depth

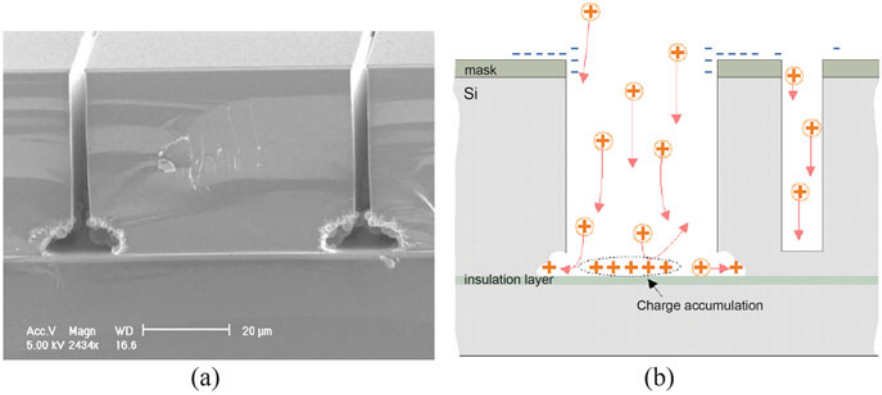


Fig. 7.25 (**a**) Notching in DRIE of SOI substrate (reprinted from [64] with permission); (**b**) mechanism of notching during etching of SOI substrate (reprinted from [67] with permission)

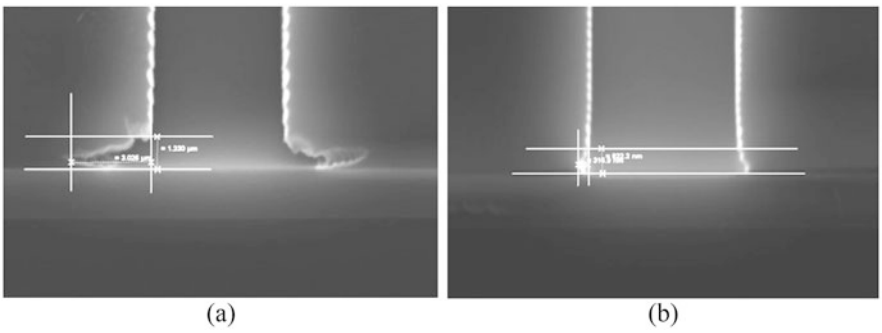


Fig. 7.26 (**a**) DRIE of trench on SOI wafer with notching; (**b**) notch-free due to low frequency pulsed biasing. (courtesy of Scott Munro at the University of Alberta's NanoFab)

(the thickness of device layer). There should be no microloading effect or ARDE. However, it is found that big features are continuously etched laterally at the foot of trenches even though the etching has stopped in depth direction. Small features can also have notches if they are over-etched. The notching, or called "footing," is believed to be caused by positive charging at the bottom insulating layer, as illustrated in Fig. 7.25b. Positive charges accumulated at the SiO_2 layer create a local positive field which deflects the incoming ions toward sides of the surface, causing the lateral etching [66]. There is nothing one can do to prevent the notching, unless the positive charge can be conducted away [67]. The most effective measure to prevent notching is to apply a low frequency pulsed bias power to the substrate stage to give the insulating layer a chance to dissipate the accumulated charge during the bias-off period [68]. Fig. 7.26 shows the etched trench on a SOI wafer by Bosch process at 13.56 MHz frequency (Fig. 7.26a) and 380KHz frequency pulsed biasing process (Fig. 7.26b). Notch-free DRIE of SOI wafer has been achieved [69].

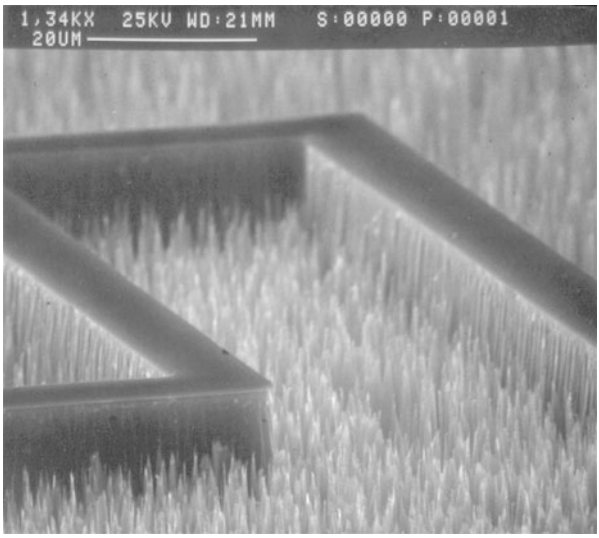


Fig. 7.27 SEM image of "black silicon"

7.4.3.4 Micromasking Effect

DRIE of silicon often encounters the "black silicon" phenomenon. The etched surface looks black color instead of shiny bright color. Close observation of the black surface finds "grasslike" structures, as the SEM image shown in Fig. 7.27, which absorbs most light with little reflection. Such grassy silicon is formed by micromasking effect [70]. During DRIE, masking layer whether photoresist or oxide can be etched due to sputtering. Droplets of masking material land on the silicon surface, which prevent the silicon underneath to be etched. So the spikes of silicon like grass are formed. By adding O_2 or CHF_3, as well as slightly tuning either bias power or substrate temperature, the spikes can be removed. The basic idea is to adjust the etching gas constituents and etching condition to increase the isotropic etching, so that these "grasses" can be cut off. On the other hand, the black silicon is a perfect antireflection surface. Its reflectivity can be less than 0.001% [71], which is a valued property for a wide range of applications such as solar cells, image sensors, photodetectors and thermal imaging cameras, and even antibacterial surface [72].

7.4.3.5 Ion Induced Damage Effect

Ion sputtering during RIE can introduce damage to the substrate. Though the energy of ions is much lower than focused ion beam sputtering, in the range of 50–500 eV, and the penetration depth is only a few nanometers, it can change the surface conductivity and the mobility of charge carriers with depth up to 100 nm [22]. The much deeper damage range is due to the ion channeling effect, e.g., ions can travel much larger distance in a crystalline material if the incident ions happened to be in

the path of low index lattice (Sect. 4.4 of Chap. 4). It is generally believed that this damage is due to the introduction of traps, vacancies, and interstitial atoms into the crystalline lattice. For single crystalline materials such as silicon, the ion induced damage can always be reduced or eliminated by post etching annealing. However, high annealing temperature can cause decomposition of compound semiconductors such as GaAs. Reducing ion energy is the most effective way of reducing the damage. For example, it is found that when RF power increased from 20 W to 40 W, the carrier mobility and concentration at the etched surface decreased by 13.8% and 29.2%, respectively [73]. Use of ICP etchers can certainly help to control the ion bombardment energy, because the substrate stage bias voltage can be independently controlled. However, detailed study found that even with ICP, one should avoid generating atomic ions in the plasma because etching with atomic ions such as Si^+ showed more damage than with molecular ions such as $(SiCl)^+$ [22]. On the other hand, reducing RF power may have to involve changing other conditions in order to maintain necessary etch rate, selectivity, and anisotropy. As a general rule, the use of highly reactive chemistries and molecular constituents within the plasma is best for reducing the effects of damage.

7.5 Atomic Layer Etching (ALE)

Previously introduced RIE has several processes taking place simultaneously as illustrated in Fig. 7.12. Atomic layer etching (ALE) only allows to remove one atomic layer a time. This self-limited surface reaction is realized by a cyclic absorption and reaction process which is illustrated in Fig. 7.28. The process starts with a fresh substrate (Fig. 7.28a). In the first reaction step, a precursor gas is introduced into the reactor for a short time (Fig. 7.28b). The chamber is then exhausted to remove the chemical reactants (Fig. 7.28c), so that only a monolayer of precursor molecules is absorbed on the substrate (Fig. 7.28d). Afterward, a beam of low energy ions is introduced to bombard the precursor covered surface, which induces a chemical reaction with the substrate atoms to form volatile compounds (Fig. 7.28e). These volatile compounds are pumped away (Fig. 7.28f) in second purge, resulting in a monolayer of substrate to be etched (Fig. 7.28g) [74].

To enable successful ALE, the following requirements should be met: (1) the precursor is chosen so that upon reaction with substrate atoms volatile products can be formed; (2) the precursor does not spontaneously etch the substrate; (3) energies for ion or neutral beam bombardment are typically limited to 100 eV or lower to minimize physical sputtering and surface damage; and (4) the timing of purge should be precisely controlled so that the substrate thickness only loses about one monolayer per cycle. The self-limited surface reaction is ensured by absorption of monolayer reactant molecules and by minimizing both spontaneous chemical etching and physical sputtering. In previously described RIE, SF_6 is the dominant etchant chemical compound for silicon etching. However, spontaneous etching takes place at room temperature for the F/Si system, which can only be suppressed at cryogenic

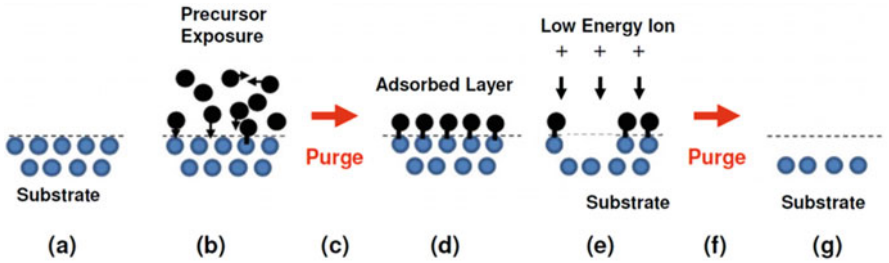


Fig. 7.28 Process sequence of ALE: (**a**) fresh substrate; (**b**) introducing precursor gas; (**c**) purge the precursor gas; (**d**) a monolayer absorption of reactant on substrate; (**e**) low energy ions bombarding the absorbed monolayer; (**f**) volatile compound is purged; (**g**) an atomic layer of substrate is etched

temperatures. For this reason, ALE of silicon adopts the Cl_2/Si system because the reaction of monolayer absorption of Cl_2 on silicon only takes place with the aid of low energy ion (Ar^+ at 20 eV) bombardment at room temperature [75]. Although the surface reaction of Cl_2/Si system can also be thermally initiated, it requires significantly high temperatures (>600 °C). Use of an Ar plasma can lower the temperature requirement to below 200 °C [76]. The advantage of thermal ALE is that the removal of material is isotropic, which implies that it can be used to thin down materials in almost any geometry.

ALE was first reported in 1990 [77], which was initially developed for GaAs etching to avoid damages caused by ion bombardment in conventional RIE. In the three decades of research and development, over two dozen materials have been shown capable of being processed with this technique [76]. It remained only as a topic of scientific interest because the etch rate is too low to draw industrial attention. This has changed in recent years as the mainstream semiconductor fabrication processes moved to sub-10 nm nodes [78]. At such length scales, all of lithography, material deposition, and material removal need to be capable of a few nanometer resolution, and this is the reason for the semiconductor industry's interest in atomic layer processes. As the atomic layer deposition (ALD) has been widely used for material deposition in the latest IC manufacturing, ALE is being driven by the semiconductor industry as an alternative to conventional RIE and is viewed as an essential counterpart to ALD. Because lateral shrinkage of device features also brings about vertical diminution of layer thicknesses, ALD and ALE are the only two processes that are commercially available for such purposes. Apart from atomic precision removal of materials, ALE has other advantages over conventional RIE, including no or very little ion bombardment-related damage to the critical gate region and significantly reduced process variability (both across the wafer and die-to-die).

7.6 Ion Milling

Etching at sub-100 nm dimension is also possible by ion milling. Ion milling is a pure physical ion sputtering process. Ion milling is needed because not all materials can be etched by RIE. Only those materials which chemically react with reactive ions and form volatile compounds can be etched by RIE. For etching other materials, ion milling is the alternative.

Argon gas is commonly used to generate gaseous ions in ion milling. As argon is an inert gas, argon ions do not react chemically with any sample materials. There are different physical reactions between Ar + and sample surface, depending on the Ar + energy. If Ar + energy is less than 10 eV, surface absorption or surface damage is the dominant process. Ion damage to sample surface confines to only a few atomic layers. If Ar + energy is over 10 keV, ions can penetrate deep inside sample and become ion implantation. Ion milling only occurs at Ar + energy between 10 and 5000 eV. It has been shown in previous sections that there is always ion sputtering taking place in RIE. However, ions in a RIE system are generated in the etching chamber and only accelerated by the negative field near the sample surface (Fig. 7.11). The sputtering efficiency is low, and material removal by pure ion sputtering is not the dominant etching mechanism.

Figure 7.29 is the schematic of a typical ion milling system. The ion source is a plasma source which has been described in Sect. 4.2.2 of Chap. 4. In the ion source, electrons emitted from a thermionic cathode are accelerated by the field between cathode and anode. Argon molecules are ionized by energized electrons, and plasma is formed in the ion source chamber. The argon ions are then extracted from ion source chamber in the form of a broad beam and accelerated toward the sample in the sputtering chamber. Therefore, ion milling is also called ion beam milling or ion beam etching (IBE). It is different from the focused ion beam that has been introduced in Chap. 4, but the sputtering mechanism is the same. Because Ar + is chemically inert to sample materials, the etch rate by ion milling only depends on the sputtering yield of a particular sample material. Typical ion milling rate is 100~3000 A/min at 1000 eV Ar ion energy and 1 mA/cm^2 ion current density. For sputtering an insulating material, an electron gun, which is a hot filament as shown in Fig. 7.29, is employed to provide flood exposure of electrons for neutralizing positive charge accumulated at the sample surface.

In principle, ion milling can etch any materials. Its independence of material types makes the etching technique a good supplement to RIE and other etching techniques. It opens up the possibility to fabricate nanostructures in a much broader range of materials other than those commonly used in microelectronic industry. For example, many magnetic materials can be etched by ion milling [79, 80]. However, one must be aware of the following issues when using ion milling for nanoscale pattern transfer:

(1) Physical sputtering by Ar ions is basically nonselective to materials. The etch selectivity only depends on the difference of sputtering yield for different materials, which are relatively small. Table 7.5 shows the ion milling rates for

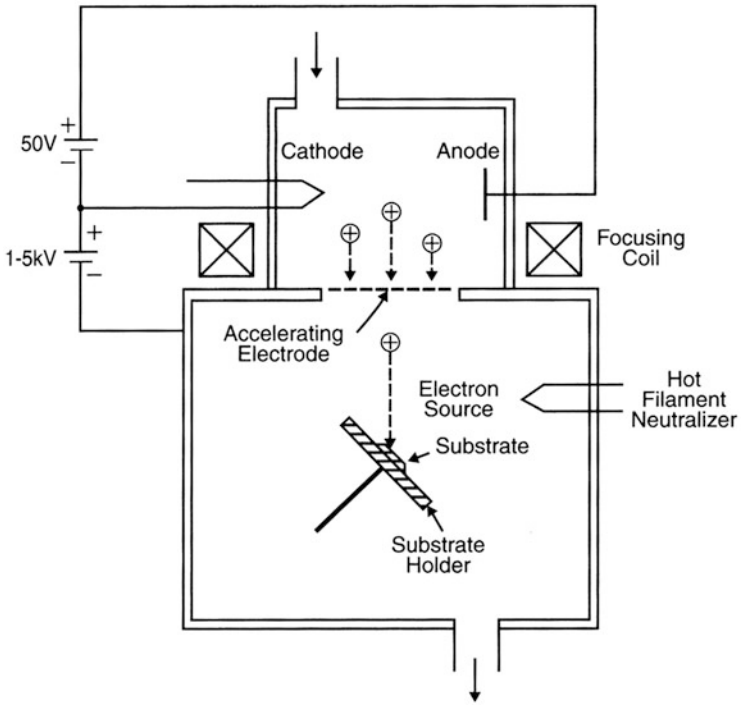


Fig. 7.29 Schematic of ion beam milling system

Table 7.5 Etching rates of some materials by ion beam sputtering

Materials	Etch rate (A°/min)
Gold	1000
Copper	700
Photoresist (AZ-1350)	200
Nichrome (NiCr)	170
Alumina (Al_2O_3)	90

some commonly used material in electronic industry. It is apparent that photoresist, which is the basic masking material in most RIE processes, is only marginally suitable as mask for ion milling of gold film, but not for nichrome or alumina.

For some materials, the selectivity of photoresist to the etched material can be improved by introducing certain chemical gases, so that the ion milling becomes partially reactive. Figure 7.30 has compared ion beam milling of hafnium oxide (HfO_2) as the function of ion energy with and without adding CHF_3 gas [81]. Without the participation of CHF_3 in the milling process (Ar only), the etching rate of photoresist (PR) is higher than that of HfO_2. By mixing two parts of CHF_3 with

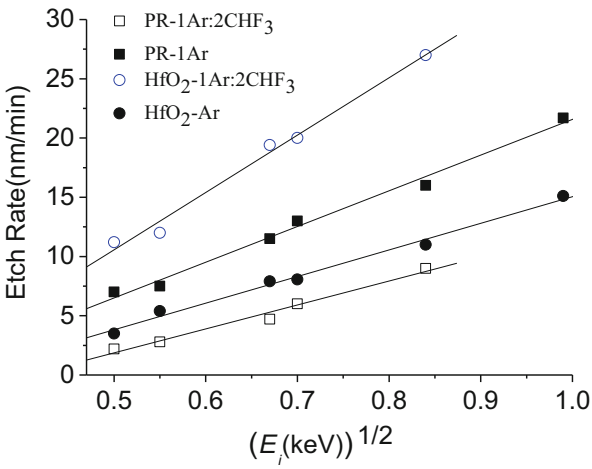


Fig. 7.30 Etch rates of hafnium oxide (HfO_2) and photoresist (PR) as the function of square root ion energy $(E_i)^{1/2}$

one part of Ar, the etch rate of HfO_2 is much enhanced, whereas the etch rate of photoresist is reduced. In this way, the photoresist serves as an ion milling mask better. It should be noted that the reactive ion milling is by no means a true sense of RIE because the etched product is still nonvolatile.

For sub-100 nm pattern transfer by ion milling, e-beam resists have to be used as the masking layer because of the use of e-beam lithography for patterning. Though PMMA is the popular high resolution e-beam resist, it is not a good masking material for ion milling [80]. The negative tone inorganic resist HSQ works better. Pattern features as small as 30 nm have been made with HSQ by e-beam lithography and served as ion milling mask to etch gold nanostructures [82]. The gold nanostructures produced by combining negative tone HSQ mask and ion milling are equivalent to those fabricated by lift-off (see next chapter) in positive resist mask but have much smoother edges than those done by lift-off. The thickness of HSQ masking layer has to be similar to the thickness of metal layer to be etched, because the selectivity of HSQ in ion milling of gold is still low [82].

In general, the etch depth by ion milling is limited because of the fast consumption of mask material. As mask thickness is reduced toward the end of etching, the final etched feature may deviate from its original design. The most pronounced profile distortion encountered in ion milling pattern transfer is faceting. Faceting is a phenomenon that top corners of a masking pattern are eroded faster than its planar surface, as illustrated in Fig. 7.31. Figure 7.31a is the initial mask pattern structure before ion milling. Because of the fast sputtering at the pattern corner, the mask pattern becomes sloped in the course of ion milling, so does the underlying etched pattern structure as shown in Fig. 7.31b. The faceting is most severe for polymer resist mask because much thick masking layer has to be used. To reduce the faceting problem, the masking layer should be made very thin. Hard metallic masking materials can be used, such as tungsten, which has been proven to alleviate the

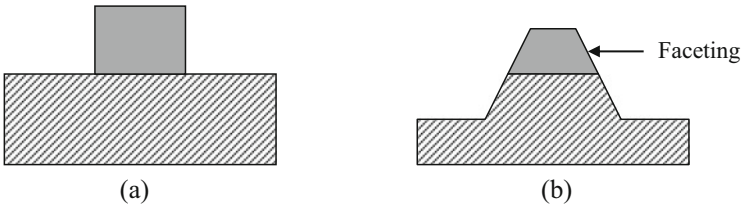


Fig. 7.31 Faceting in ion milling: (**a**) initial mask pattern profile; (**b**) faceting of etched profile caused by ion sputtering

faceting problem [83]. However, patterning the hard mask needs an additional fabrications step, which inevitably complicates the pattern transfer process.

(2) Ion milling is a physical etching process and does not form volatile compounds. The sputtered material can redeposit everywhere, particularly onto adjacent patterns [81]. By rotating while tilting the sample during sputtering, it helps to clean up effectively the redeposition of sputtered debris on the structure sidewalls [80]. A gentle wet acidic etch after ion milling can also help to clean up the redeposition, though the acid that removes the redeposited material can harm the etched structures as well. To overcome the dilemma, some reactive gases can be introduced into the ion milling process, which may produce some nonvolatile compounds dissolvable in acid but without attacking the pattern structures. For example, by ion milling of HfO_2 with Ar/CHF_3 gas mixture, the reaction product is hafnium fluoride which can be dissolved in HCl acid, but the HCl does not react with HfO_2 [81].

References

1. Biebuyck, H.A., et al.: Lithography beyond light: microcontact printing with monolayer resists. IBM J Research & Development. **41**(1/2), 159 (1997)
2. Xia, Y., Whitesides, G.M.: Soft lithography. Angew. Chem. Int. Ed. **37**, 550–575 (1998)
3. Kendall, D.L., Shoultz, R.A.: Wet chemical etching of silicon and SiO2 and ten challenges for micromachiners. In: Rai-Coudhury, P. (ed.) Handbook of Microlithography, Micromachining and Microfabrication. SPIE Optical Engineering Press (1997)
4. Kendall, D.L., Shoultz, R.A.: Wet chemical etching of silicon and SiO2 and ten challenges for micromachiners. In: Rai-Coudhury, P. (ed.) Handbook of Microlithography, Micromachining and Microfabrication. SPIE Optical Engineering Press & IEE (1997)
5. IntelliSuite™ from IntelliSense Software Corp. https://www.intellisense.com/
6. Frasca, S., et al.: The Michelangelo step: removing scalloping and tapering effects in high aspect ratio through silicon vias. Sci. Rep. **11**, 3997 (2021)
7. Lai, J.M., Chieng, W.H., Huang, Y.C.: Precision alignment of mask etching with respect to crystal orientation. J. Micromech. Microeng. **8**, 327 (1998)
8. Sugimura, H., Nakagiri, N.: Fabrication of silicon nanostructures through scanning probe anodization followed by chemical etching. Nanotechnology. **6**, 29–33 (1995)
9. Zhang, Y.Y., et al.: Fabrication of silicon-based multilevel nanostructures via scanning probe oxidation and anisotropic wet etch. Nanotechnology. **16**, 422–428 (2005)

10. Li, B., et al.: Fabrication and characterization of patterned single-crystal silicon Nanolines. Nano Lett. **8**(1), 92–98 (2008)
11. Wieser, U., et al.: Nanoscale patterning of Si/SiGe heterostructures by electron-beam lithography and selective wet-chemical etching. Semicond. Sci. Technol. **15**, 862–867 (2000)
12. Li, X., Bohn, P.W.: Metal-assisted chemical etching in HF/H2O2 produces porous silicon. Appl. Phys. Lett. **77**(16), 2572–2574 (2000)
13. Huang, Z., et al.: Metal-assisted chemical etching of silicon: a review. Adv. Mater. **23**, 285–308 (2011)
14. Huang, Z., Fang, H., Zhu, J.: Fabrication of silicon nanowire arrays with controlled diameter, length, and density. Adv. Mater. **19**, 744 (2007)
15. Chang, S.-W., et al.: Densely packed arrays of ultra-high-aspect-ratio silicon nanowires fabricated using block-copolymer lithography and metal-assisted etching. Adv. Funct. Mater. **19**, 2495–2500 (2009)
16. Tiberio, R.C., et al.: Vertical directionality-controlled metal-assisted chemical etching for ultrahigh aspect ratio nanoscale structures. J. Vac. Sci. Technol. **32**(6), 06FI01-1 (2014)
17. Ravi, P.: Srivastava and dahl-young Khang, structuring of Si into multiple scales by metal-assisted chemical etching. Adv. Mater. **33**, 2005932 (2021)
18. Romano, L., et al.: Metal assisted chemical etching of silicon in the gas phase: a nanofabrication platform for X-ray optics. Nanoscale Horiz. **5**, 869 (2020)
19. Donnelly, V.M., Kornblit, A.: Plasma etching: yesterday, today, and tomorrow. J. Vac. Sci. Technol. A. **31**(5), 050825 (2013)
20. Laermer, F.: Dry etching. In: Zappe, H.P., et al. (eds.) Comprehensive Microsystems, pp. 217–233. Elsevier (2008)
21. Hitchon, W.N.G.: Plsama processes for semiconductor fabrication. In: Ahmad, H. (ed.) Cambridge Studies in Semiconductor Physics and Microelectronic Engineering. Cambridge University Press (1999)
22. Wilkinson, C.D.W., Rahman, M.: Dry etching and sputtering. Phil. Trans. R. Soc. Lond. **A362**, 125–138 (2003)
23. Wang, X., et al.: Fabrication of nanoimprint template in Si with high etch rate by non-switch DRIE proces. Microelectron. Eng. **85**(5–6), 1015–1017 (2008)
24. Peters, L.: Plasma etch chemistry: the untold story. Semicond. Int. **15**(6), 66 (1992)
25. Tian, W.-C., Weigold, J.W., Pang, S.W.: Comparison of Cl_2 and F-based dry etching for high aspect ratio Si microstructures etched with an inductively coupled plasma source. J. Vac. Sci. Technol. **B18**(4), 1890–1896 (2000)
26. Cardinaud, C., Peignon, M.-C., Tessier, P.-Y.: Plasma etching: principles, mechanisms, application to micro- and nano-technologies. Appl. Surf. Sci. **164**, 72–83 (2000)
27. Tao, J., et al.: A systematic study of dry etch process for profile control of silicon tips. Microelectron. Eng. **78–79**, 147–151 (2005)
28. Grigoropoulos, S., et al.: Highly anisotropic silicon reactive ion etching for nanofabrication using mixtures of SF6/CHF3 gases. J. Vac. Sci. Technol. **B15**(3), 640 (1997)
29. Chen, Y., et al.: Nanoimprint lithography for planar chiral photonic meta-materials. Microelectron. Eng. **78–79**, 612–617 (2005)
30. Schutz, R.J.: Reactive plasma etching. In: Sze, S.M. (ed.) VLSI Technology. McGraw-Hill Book Co. (1988)
31. Chen, H., Fu, C.: An investigation into the characteristics of deep reactive ion etching of quartz using SU-8 as a mask. J. Micromech. Microeng. **18**, 105001 (2008)
32. Lee, J.W., et al.: Inductively coupled plasma etching of III-V semiconductors in Cl_2-based chemistries. Mater. Sci. Semicond. Process. **1**, 65–73 (1998)
33. Osipov, A.A., et al.: Dry etching of silicon carbide in ICP with high anisotropy and etching rate. IOP Conf. Series. Mater. Sci. Eng. **862**, 022039 (2020)
34. Xia, Y., et al.: Inductively coupled plasma etching of bulk tungsten for MEMS applications. Sensors Actuators A Phys. **345**, 113825 (2022)
35. ICP etcher vendor: Oxford Plasma Technology. https://plasma.oxinst.com/

36. ICP etcher vendor: Plasma-Therm. https://www.plasmatherm.com/
37. ICP etcher vendor: SENTECH. https://www.sentech.com/
38. ICP etcher vendor: SPTS Technologies. https://www.spts.com/
39. Gambino, J.P., Adderly, S.A., Knickerbocker, J.U.: An overview of through-silicon-via technology and manufacturing challenges. Microelectron. Eng. **135**, 73–106 (2015)
40. Petti, C.: 3D memory: etch is the new Litho. Proc. of SPIE. **10589**, 1058904 (2018)
41. Laermer, F., Urban, A.: MEMS at Bosch-Si plasma etch success story, history, applications, and products. Plasma Process. Polym. **16**, 1800207 (2019)
42. Ayon, A.A., Bayt, R.L., Breuer, K.S.: Deep reactive ion etching: a promising technology for micro- and nanosatellites. Smart Mater. Struct. **10**, 1135–1144 (2001)
43. Choi, C.-H., Kim, C.-J.: Fabrication of silicon nanostructures with various sidewall profiles and sharp tips. In: Proceedings of Transducers'05: the 13th International Conference on Solid-State Sensors, Actuators and Microsystems. IEEE, Seoul (2005)
44. Nilsson, D., Jensen, S., Menon, A.: Fabrication of silicon molds for polymer optics. J. Micromech. Microeng. **13**, S57–S61 (2003)
45. Agarwal, R., Samson, S., Bhansali, S.: Fabrication of vertical mirrors using plasma etch and KOH:IPA polishing. J. Micromech. Microeng. **17**, 26–35 (2007)
46. Teo, S.H.G., et al.: High resolution and aspect ratio two-dimensional photonic band-gap crystal. J. Vac. Sci. Technol. **B 22**(6), 2540 (2004)
47. Kawata, H., Yasuda, M., Hirai, Y.: Fabrication of Si mold with smooth side wall by new plasma etching process. Microelectron. Eng. **84**, 1140–1143 (2007)
48. Welch, C.C., et al.: Silicon etch process options for micro- and nanotechnology using inductively coupled plasmas. Microelectron. Eng. **83**, 1170–1173 (2006)
49. Mirza, M.M., et al.: Nanofabrication of high aspect ratio (∼50:1) sub-10 nm silicon nanowires using inductively coupled plasma etching. J. Vac. Sci. Technol. B. **30**(6), 06FF02-1 (2012)
50. Alsolami, A., et al.: Improvement of deep reactive ion etching process for motional resistance reduction of Capacitively transduced vibrating resonators. IEEE Sensors Letters. **2**(3), 1–4 (2017)
51. Parasuraman, J., et al.: Deep reactive ion etching of sub-micrometer trenches with ultra high aspect ratio. Microelectron. Eng. **113**, 35–39 (2014)
52. Walker, M.: Comparison of Bosch and cryogenic processes for patterning high aspect ratio features in silicon. SPIE. **4407**, 89 (2001)
53. Henry, M.D.E.A.: Techniques of cryogenic reactive ion etching in silicon for fabrication of sensors. J. Vac. Sci. Technol. A. **27**, 1211–1216 (2009)
54. Tachi, S., Tsujimoto, K., Okudaira, S.: Low-temperature reactive ion etching and microwave plasma etching of silicon. Appl. Phys. Lett. **52**(8), 616–618 (1988)
55. Gottscho, R.A., Jurgensen, C.W.: Microscopic uniformity in plasma etching. J. Vac. Sci. Technol. **B10**(5), 2133 (1992)
56. Karttunen, J., Kiihamäki, J., Franssila, S.: Loading effects in deep silicon etching. Proc. SPIE. **4174**, 90–97 (2000)
57. Huff, M.: Recent advances in reactive ion etching and applications of high-aspect-ratio microfabrication. Micromachines. **12**, 991 (2021)
58. Chen, A., et al.: Fabrication of sub-100 nm patterns in SiO_2 templates by electron-beam lithography for the growth of periodic III–V semiconductor nanostructures. Nanotechnology. **17**, 3903–3908 (2006)
59. Feurprier, Y., et al.: Microloading effect in ultrafine SiO2 hole/trench etching. J. Vac. Sci. Technol. **A17**(4), 1556 (1999)
60. Dalton, T.J., et al.: Microtrench formation in polysilicon plasma etching over thin gate oxide. J. Electrochem. Soc. **140**(8), 2395 (1993)
61. Hoekstra, J.R., Kushner, M.J.: Microtrenching resulting from specular reflection during chlorine etching of silicon. J. Vac. Sci. Technol. **B16**(4), 2102 (1998)

62. Schaepkens, M., Oehrlein, G.S.: Asymmetric microtrenching during inductively coupled plasma oxide etching in the presence of a weak magnetic field. Appl. Phys. Lett. **72**(11), 1293 (1998)
63. Watanabe, M., Shaw, D.M., Collins, G.J.: Reduction of microtrenching and Island formation in oxide plasma etching by employing electron beam charge neutralization. Appl. Phys. Lett. **79**(22), 2698 (2001)
64. Kim, J., et al.: Robust SOI process without footing and its application to ultra high-performance microgyroscop. Sensors Actuators A Phys. **114**, 236–243 (2004)
65. Lasky, J.B.: Wafer bonding for silicon-on-insulator technology. Appl. Phys. Lett. **48**, 78–80 (1986)
66. Hwang, G.S., Giapis, K.P.: On the origin of the notching effect during etching in uniform high density plasmas. J. Vac. Sci. Technol. **B15**(1), 70 (1997)
67. Kim, C.-H., Kim, Y.-K.: Prevention method of a notching caused by surface charging in silicon reactive ion etching. J. Micromech. Microeng. **15**, 358–361 (2005)
68. Laermer, F., Urban, A.: Challenges, developments and applications of silicon deep reactive ion etching. Microelectron. Eng. **67-68**, 349–355 (2003)
69. Munro, S.: SOI Wafer Notch Reduction Using the Low Frequency Pulsing Option on the STS ICP-RIE. University of Alberta, NanoFab (2009)
70. Jansen, H., et al.: The black silicon method: a universal method for determining the parameter setting of a fluorine-based reactive ion etcher in deep silicon trench etching with profile control. J. Micromech. Microeng. **5**(2), 115–120 (1995)
71. Liu, T., Dickensheets, D.L.: Black silicon integrated aperture. J. Micro/Nanolith. MEMS MOEMS. **16**(4), 045501 (2017)
72. Zheng, F., et al.: Recent Progress of black silicon: from fabrications to applications. Nano. **11**, 41 (2021)
73. Li, X., et al.: A low damage RIE process for the fabrication of compound semiconductor based transistors with sub-100 nm tungsten gates. Microelectron. Eng. **83**, 1159–1162 (2006)
74. Oehrlein, G.S., Metzler, D., Li, C.: Atomic layer etching at the tipping point: an overview. ECS journal of solid state. Sci. Technol. **4**(6), N5041–N5053 (2015)
75. Matsuura, T., et al.: Self-limited layer-by-layer etching of Si by alternated chlorine adsorption and Ar+ ion irradiation. Appl. Phys. Lett. **63**(20), 2803–2805 (1993)
76. Rahman, F.: Atomic layer processes for material growth and etching—a review. IEEE Trans. Semicon. Manufacturing. **34**(4), 500–512 (2021)
77. Meguro, T., et al.: Digital etching of GaAs: new approach of dry etching to atomic ordered processing. Appl. Phys. Lett. **56**(16), 1552–1554 (1990)
78. Kanarik, K.J., et al.: Overview of atomic layer etching in the semi-conductor industry. J. Vac. Sci. Technol. A. **33**, 020802 (2015)
79. Maat, S., et al.: Ultrathin CoPt-pinned current perpendicular to the plane spin valves. J. Appl. Phys. **98**(11), 113907 (2005)
80. Moneck, M.T., et al.: Fabrication of flyable perpendicular discrete track media. IEEE Trans. Magn. **43**(6), 2127–2212 (2007)
81. Wang, X., et al.: Reactive ion beam etching of HfO_2 film and removal of sidewall Redeposition. J. Vac. Sci. Technol. **A24**(4), 1067 (2006)
82. Stade, F., et al.: Fabrication of metallic nanostructures for investigating plasmon-induced field enhancement. Microelectron. Eng. **84**, 1589–1592 (2007)
83. Walsh, M.E., et al.: Optimization of a lithographic and ion beam etching process for nanostructuring magnetoresistive thin film stacks. J. Vac. Sci. Technol. **B18**(6), 3539 (2000)

Chapter 8
Nanoscale Pattern Transfer by Deposition



8.1 Introduction

Although subtractive etching is the most used pattern transfer process, not all materials can be etched either by wet chemical or dry plasma. Metals in particular are difficult to etch by reactive plasma process except only a few. Though wet acids can etch metals, the isotropic nature of wet etch means it is difficult to obtain high resolution as well as high aspect ratio metallic structures. Pure physical sputtering such as the ion milling introduced in Chap. 7 is too slow and has stringent requirements on the mask. Therefore, etching alone cannot perform all the pattern transfer tasks in nanofabrication. The other option is to transfer resist patterns by deposition. It is an additive process, i.e., adding materials through the resist pattern openings. The emphasis in this chapter is the additive pattern formation through a mask, as opposed to direct additive pattern formations introduced in previous chapters, such as patterned depositions by electron beam or focused ion beam or scanning probes. For masked pattern transfer where the mask is a polymer resist pattern created by lithography, whether using optical or electron beam or nanoimprint, patterns of deposited materials can be realized by liftoff as shown in Fig. 8.1a or by plating as shown in Fig. 8.1b. Other additive techniques include pattern transfer by shadow masks or by direct printing.

8.2 Thin Film Deposition

The additive pattern transfer process involves depositing thin films into the opening of a polymer resist mask. A thin film can be deposited by physical vapor deposition (PVD), chemical vapor deposition (CVD), atomic layer deposition (ALD), molecule beam epitaxy (MBE), oxidation, plating, spin or spray coating, blade coating, or printing. PVD and CVD are the two most commonly used processes [1]. PVD

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_8

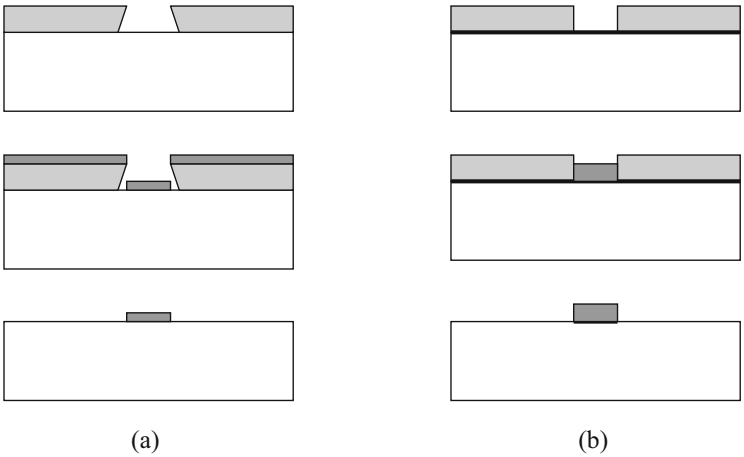


Fig. 8.1 Schematic of additive pattern transfer processes: (**a**) liftoff, (**b**) electroplating

includes thermal evaporation deposition and plasma sputtering deposition. The thermal evaporation can further be divided into resistive heating evaporation and electron beam bombardment heating evaporation. The plasma sputtering is differentiated according to how the plasma is generated. It can be a DC plasma sputtering, magnetron plasma sputtering, or radio frequency (RF) plasma sputtering. CVD includes low-pressure CVD (LPCVD), plasma enhanced CVD (PECVD), atmospheric pressure CVD (APCVD), and metal organic CVD (MOCVD). ALD is a pure chemical process by which single layer atoms are deposited by self-limited surface chemical reaction [2]. The main features of PVD, CVD, and ALD thin film deposition techniques are summarized in Table 8.1.

One of the immediate concerns, when thin film deposition process is used for additive pattern transfer, is the substrate temperature. As described earlier, polymer resist which is the masking material for the pattern transfer cannot withstand temperature higher than its glass transition temperature (T_g). As it is known from the previous chapters, at above T_g temperature, polymers become soften and flowable. Distortions become inevitable particularly for small dimension and high aspect ratio patterns. The requirement for low substrate temperature leaves only the PVD as the suitable thin film deposition technique for additive pattern transfer.

Another requirement is the directionality of deposition. Additive pattern transfer by liftoff needs the thin film deposition to have poor step coverage, so that the film deposited onto the resist masking layer can be easily removed by dissolving the polymer resist. Good step coverage will connect the film between top and bottom of polymer resist patterns, making the liftoff difficult if not impossible. Figure 8.2 illustrates the effect of step coverage on the final outcome of liftoff. With a poor step coverage, the resist layer can be easily removed by solvents together with deposited film on top of it, as shown in Fig. 8.2a, whereas a good step coverage prevents solvent from penetrating into the resist layer, making the complete removal of resist

Table 8.1 Characteristics of some major thin film deposition techniques

Techniques	Materials	Film uniformity	Film density	Film grain size	Deposition rate	Substrate temperature	Directionality
Resistive heating evaporation	Low melting point metal or other materials	Poor	Poor	10–100 nm	1–20 Å /s	50–100 °C	Good
Electron beam bombardment evaporation	High melting point metals and dielectrics	Poor	Poor	10–100 nm	10–100 Å /s	50–100 °C	Good
Plasma sputtering	Metals and dielectrics	Very good	Good	~10 nm	Metal: ~100 Å /s Dielectric:1–10 Å /s	~200 °C	Moderate
PECVD	Dielectrics	Good	Good	10–100 nm	10–100 Å /s	200–300 °C	Moderate
LPCVD	Dielectrics	Very good	Excellent	1–10 nm	10–100 Å /s	600–1200 °C	Poor
ALD	Metals and dielectrics	Excellent	Excellent	<1 nm	~1 Å/cycle	25–300 °C	Very poor

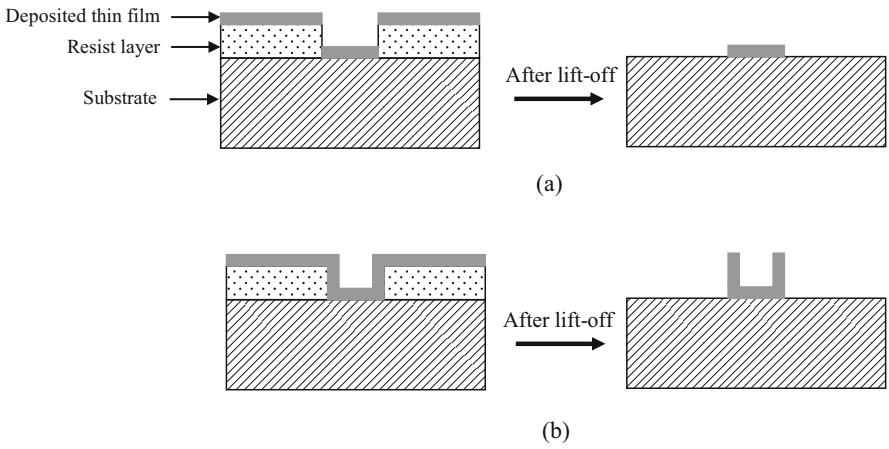


Fig. 8.2 Liftoff with (**a**) good directionality and (**b**) poor directionality of thin film deposition

as well as the deposited film difficult. Even though the resist is removed, residues of deposited film can remain, as shown in Fig. 8.2b. The poor step coverage is ensured by good directionality of deposition. As seen in Table 8.1, this leaves only the thermal evaporation deposition as the most suitable thin film deposition technique for liftoff pattern transfer. Thin films deposited by CVD or ALD techniques are primarily for subtractive pattern transfer process, i.e., the films have already been deposited on a substrate before the polymer resist layer is lithographically patterned.

Because of its high directionality, the deposition rate of thin film by thermal evaporation depends on the distance and relative position between substrate and evaporation source. It follows the Knudsen's cosine law, as expressed in Eq. 8.1

$$R \propto \frac{\cos\theta \cos\phi}{r^2}\left(P_s(T) - P_c\right) \tag{8.1}$$

In the above equation, R is the deposition rate which is the function of evaporation source to the sample distance r, the angle (θ) of sample to the axis of evaporation source, and the orientation of sample surface ϕ. The source position relative to a sample is schematically indicated in Fig. 8.3a. The rate also depends on the vapor pressure of source material P_s which is the function of heating temperature and the pressure in the evaporation chamber P_c.

It is apparent that a flat sample placed at certain distance from the evaporation source will have nonuniform film thickness at different locations of the sample. Assuming the film thickness at the center of sample (on the axis of evaporation source) is t_0, the thickness at an off-center location with distance x is t_x, the height of sample above the source is h, and the size of evaporation source is assumed to be much smaller than the source to sample distance (h), and then t_x can be worked out from Eq. 8.2, according to the Knudsen's cosine law [3].

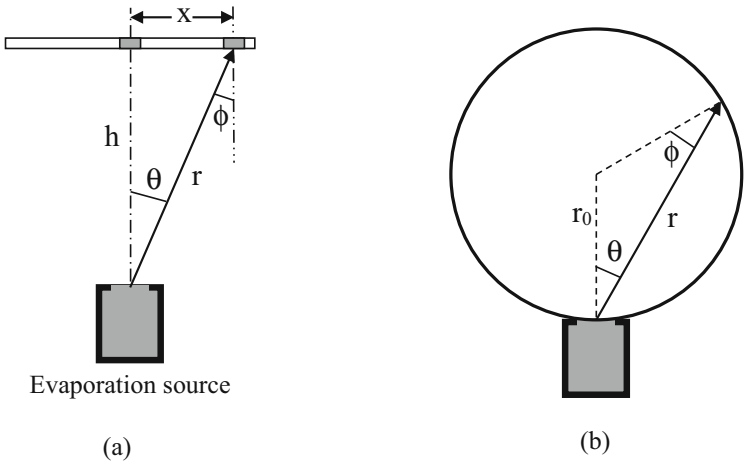


Fig. 8.3 (**a**) Thin film deposition on a flat sample by thermal evaporation; (**b**) spherical arrangement for uniform deposition by thermal evaporation

$$t_x = \frac{t_0}{\left[\left(1+\frac{x}{h}\right)^2\right]^2} \tag{8.2}$$

Eq. 8.2 indicates that film thickness at an off-center location on a flat sample is always less than at a close-to-center location. Such difference can only be reduced if the sample is further away from the evaporation source. In order to achieve uniform deposition, the sample size should be small, and the evaporation chamber should be big. All of these measures are at the price of low deposition rate and less efficient use of source material.

It is possible to obtain uniform film thickness, provided that the sample is placed on a spherical surface, as illustrated in Fig. 8.3b. In practice, samples which are usually small in size can be placed on spherically shaped sample holder. At any point of the spherical surface, the angles θ and ϕ are equal. Assuming the sphere has radius of r_0, then the distance between any point on the spherical surface and the source can be expressed as $r = 2r_0 \cos\theta = 2r_0 \cos\phi$. Replacing the distance r in Eq. 8.1, the deposition rate at the spherical surface becomes Eq. 8.3:

$$R \propto \frac{P_e - P_c}{4r_0^2} \tag{8.3}$$

Now the deposition rate only depends on the radius of sphere r_0. The film thickness should be uniform as long as the sample is placed at a position which coincides with the surface of a sphere and the evaporation source is located at the edge of this sphere as indicated in Fig. 8.3b.

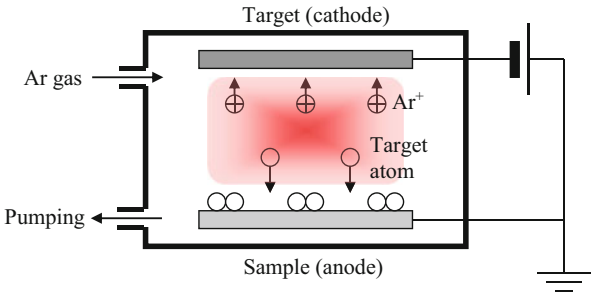


Fig. 8.4 Schematic of plasma sputtering deposition process

Another widely used PVD technique in additive pattern transfer is the plasma sputtering deposition. An important advantage of sputter deposition is that even materials with very high melting points can be easily sputtered while evaporation of these materials in a thermal evaporator is problematic or impossible. Sputtered films typically have a better adhesion on the substrate than evaporated films. A sputter coating system is normally in diode configuration, as schematically shown in Fig. 8.4. Two parallel electrodes are mounted in a vacuum chamber with a distance. One is the stage on which a sample is placed, and another is the target from which comes the deposition material. A direct current (DC) voltage of several kilovolts is applied between the electrodes, with the sample stage as anode and target as cathode. Inert gas, normally argon, is introduced into the vacuum chamber. At high electrical field, free electrons are accelerated and collide with Ar atoms which become ionized. More free electrons are available in the chamber because of the ionization. These electrons acquire energy from the electrical field between the diode electrodes, which causes more Ar atoms to be ionized. The Ar^+ ions are accelerated toward the target, sputtering off target atoms. These free atoms with initial kinetic energy fly everywhere and deposit on the sample surface as well as other surrounding surfaces. Plasma sputtering in the DC diode configuration is characterized by slow deposition rates, high voltages, and low currents. It is also inadequate for the deposition of dielectric films because of charging, arcing, and very low deposition rates. The improvement was to replace the DC power supply, shown in Fig. 8.4, with a RF power supply, generally operating at the radio frequency (RF) of 13.56 MHz. This change eliminates the charging and arcing problems when sputtering dielectrics and also leads to slightly higher deposition rates. The oscillating RF potential applied to anode and cathode results in a modification to the electron motions, which provides better energy coupling to the electrons and produces higher plasma density [4].

Further evolution from the RF sputtering system was the magnetron cathode system. This type of tool uses magnetic confinement of electrons in the plasma, which results in a much higher plasma density than in either the RF or DC diode systems. In a magnetron plasma system, the electric field is always oriented normal to the surface of the cathode. The transverse magnetic field is configured perpendicular to electric field, so that the trapped and drifting electrons are constrained to circulate many times around the cathode face, which greatly enhances the Ar ionization rate and hence the deposition rate [5].

The deposition by plasma sputtering is also characterized by the Knudsen's cosine law. However, various departures from the cosine distribution can happen, depending on the ion energy and the target material structure. The sputtering rate of low energy ions is higher at larger emission angle, and the sputtering rate of high energy ions is higher at smaller emission angle. The sputtering rate is also dependent on specific crystalline orientations of the target. The size of sputtering target is much larger than the size of an evaporation source. The sputter-off target atoms in the chamber space are frequently scattered by gas atoms before reaching the sample surface. All the above factors imply that the movement of atoms has a very broad, almost isotropic nature. This is the key reason why plasma sputtering deposition is less directional than thermal evaporation deposition. There are ways to partially improve the directionality of plasma sputtering deposition. Two techniques have been developed to filter out the angular distribution of the sputtered atoms to achieve a more normal-incident deposition process: one is long-throw deposition and another is collimated deposition [5].

Long-throw deposition is simply to increase the distance between the target and the sample holder plate. It is a mimic to the thermal evaporation deposition whereby the emission source is far apart from the sample. The reason for long throw is to decrease the pressure in the plasma chamber. It is known that to sustain a plasma, the electrode spacing L and chamber pressure P have to satisfy the condition of L•P $>$ 0.5 (cm•Torr). Then to maintain plasma while keeping low pressure, the electrode spacing has to increase. The above condition is for DC plasma. For RF and magnetron plasma, electrons acquire extra energy from extended circling. The plasma can be sustained at much lower pressure than in a DC system. At 25 cm throw distance, the working pressure in a magnetron system can be in the low 0.1 mTorr range. As the mean free path of an atom in ambient gas can be expressed as Eq. 8.4, low pressure means long mean free path for target atoms.

$$\lambda \approx \frac{5 \times 10^{-3}}{P(\text{Torr})} (\text{cm}) \tag{8.4}$$

With the throw distance equal to the cathode diameter, the ballistic and unscattered atoms will have the maximum angular divergence of about ±45°. Atoms traveling at such large angles will be unable to reach the sample, which effectively increases the directionality of deposition. The drawback of long-throw deposition is some loss of deposition material because large angular atoms deposit onto the chamber walls instead of the sample.

Directional deposition can also be obtained by inserting an array of collimating tubes between the magnetron cathode and the sample [6], as schematically shown in Fig. 8.5. The collimator functions much like the sidewalls of the chamber. It collects atoms which are traveling laterally to the cathode/sample plane. These atoms are deposited on the sidewalls of the collimator cells and hence not deposited onto the sample. Geometrically, the use of a collimator does not require any significant increase in throw distance over conventional sputter deposition, thus removing the

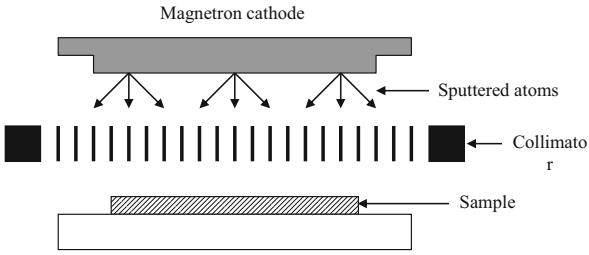


Fig. 8.5 Insertion of collimators to improve directionality in sputtering deposition

very low-pressure constraint of long-throw sputtering. The aspect ratio of the collimator can be readily changed without significantly altering its overall dimensions. For practical manufacturing applications, collimator cell diameters are of the order of 1 cm, and therefore the height of a 2:1 collimator is only 2 cm. The overall throw distance increase is then of the order of 3–4 cm. The directionality of deposition can be increased by increasing the aspect ratio of collimating tube. However, the deposition rate will be significantly reduced due to the collimation. Generally, the deposition rate is reduced by a factor of about 35% for each unit increase in the collimator aspect ratio. The reduction of deposition on the sample becomes the deposition onto the collimator tubes. Eventually this leads to either closure of the collimator holes, or peeling and flaking of thick deposits, neither of which is desirable. In practice, it is necessary to change the collimator about two to three times more often than changing of cathode, which unfortunately can significantly add to overall cost.

8.3 Pattern Transfer by Liftoff

The concept of patterning high resolution metallic structures by liftoff was proposed more than 40 years ago [7], and soon afterward sub-20 nm metal lines were made by the technique using patterned PMMA resist [8]. Liftoff has now become the primary technique for patterning metallic thin films, because for many metallic materials, there are no other means to pattern them except the liftoff. Liftoff process has been illustrated in Fig. 8.1a. A polymer resist layer is patterned by optical or electron beam lithography first. Metallic thin film is then deposited onto the patterned resist layer. Chemical solvent, normally acetone, dissolves the resist layer, which also lifts off the metallic thin film on top of the resist layer from the substrate. Only the metallic film deposited through the resist pattern opening onto the substrate remains. In this way, the resist pattern is transferred onto the substrate as a metallic pattern of reverse polarity.

The resolution capability of liftoff process is in principle the same as the resist patterning process, whether the patterning is done by optical lithography, electron beam lithography, or nanoimprinting lithography. Sub-10 nm liftoff metal patterns have been demonstrated [9]. From thin film deposition point of view, the only



(a) (b)

Fig. 8.6 Profiles of thin film coverage over a patterned resist structure with (**a**) thermal evaporation deposition and (**b**) plasma sputtering deposition

limitation on the ultimate resolution of liftoff patterns is the grain size of deposited thin films. A continuous line pattern may be broken if the grain size is comparable to the linewidth. As seen in Table 8.1, each deposition technique has its own range of grain size. Pattern transfer at sub-100 nm feature dimension requires small grain size which depends on the deposition conditions. Generally, high deposition rate, whether by thermal evaporation or plasma sputtering, will result in small grain size in the deposited film.

The key to the successful liftoff is to make sure that the liquid solvent can penetrate into the resist layer and dissolve it. This requires the metal film on top of the resist layer to disconnect with the metal film inside the resist pattern opening. The discontinuity of film is only possible if (1) the deposited film thickness is much less than the resist layer thickness and (2) the deposited film has poor step coverage over the patterned resist layer. The first condition can easily be met by using thick resist layer. For successful liftoff, the resist layer should be at least three times thicker than the deposited thin film. The second condition depends on the directionality of thin film deposition process. As described in the previous section, thermal evaporation deposition has the best directionality than any other deposition methods. Figure 8.6 compares the deposition profiles for thermal evaporation (Fig. 8.6a) and sputtering deposition (Fig. 8.6b). For this reason, whenever possible, liftoff should always be done through thermal evaporation deposition of thin film.

However, in terms of film quality, thermally evaporated film is the poorest among other methods. More importantly, thermal evaporation is not best suited for depositing alloy materials, because each constituent in an alloy may have different vapor pressure and hence different evaporation rate. The composition of deposited film may deviate from the stoichiometry of original alloy in the evaporation source. Plasma sputtering deposition, on the other hand, has more or less the same sputtering rate for different constituents in an alloy. Although different materials have different ion sputtering yield as indicated in Table 4.1 of Chap. 4, the sputtering only takes place at the target surface, which would quickly deplete the atoms of constituent of high sputtering yield, leading to decrease of its sputtering rate until all the atoms of other constituents in the surface layer are sputtered off and a new fresh layer is

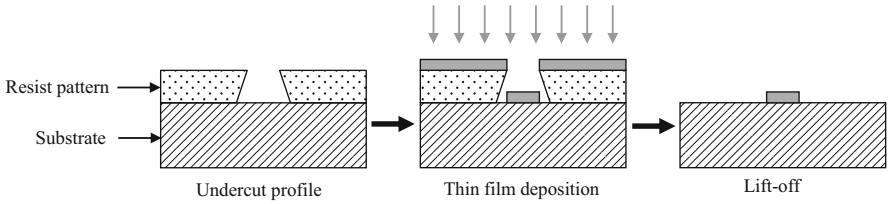


Fig. 8.7 Undercut profile of resist layer facilitates liftoff

revealed. In this way, the overall sputtering rates for all the constituents in an alloy become similar. Although thin film deposited by plasma sputtering is better, its poor directionality makes the liftoff much more difficult than using thermal evaporation deposition. In order to ensure the discontinuity of thin film over patterned polymer resist topography, it is highly desirable to create an undercut profile in the patterned resist structure, as is schematically drawn in Fig. 8.7. With an undercut profile, in addition to a thicker resist layer, the poor directionality of thin film deposition can be offset, making liftoff possible even with plasma sputtering deposition.

It is difficult to produce an undercut resist profile in optical lithography, because the absorption of photons in photoresist is always higher at top part of the photoresist layer, resulting in an overcut profile after development. In electron beam lithography, undercut resist profile is possible. As described in Chap. 3, electrons scatter in a resist layer, and electron energy deposition becomes broader as electrons going deeper in resist. This feature can be confirmed by Monte Carlo simulation of electron scattering and resist exposure. Figure 8.8a is the computer simulated resist profiles exposed with 20 keV electron beam. The resist is a 0.5-μm-thick PMMA and the pattern is single 50 nm line. The pear-shaped profile is the result of electron scattering. Three resist profiles are shown, representing three different exposure doses. The higher the exposure dose, the broader the undercut range. As described in Chap. 3, electron beam lithography at low energy does not have high resolution, particularly for dense patterns. When patterning thick resist at high resolution, especially for sub-100 nm features, high beam energy is preferred. However, high beam energy makes it difficult to produce an undercut profile in a single layer resist, as is seen in Fig. 8.8b. Increasing exposure dose makes little difference in increasing the undercut range.

The above problem can be solved by bilayer resist systems. Figure 8.9 illustrates how the bilayer resist system for liftoff works. The commonly used bilayer resist system for liftoff is the combination of PMMA and its copolymer [10]. They are both sensitive to electron exposure. High molecular weight PMMA, which is of low sensitivity to electron exposure, forms the top layer. Low molecular weight copolymer P(MMA-MAA), which is of high sensitivity to electron exposure, forms the bottom layer. They are both developed in the same developer solution after e-beam exposure. Because the bottom layer resist is more sensitive to e-beam exposure than the top layer resist, it develops faster with the same exposure dose than the top layer, which creates the undercut profile.

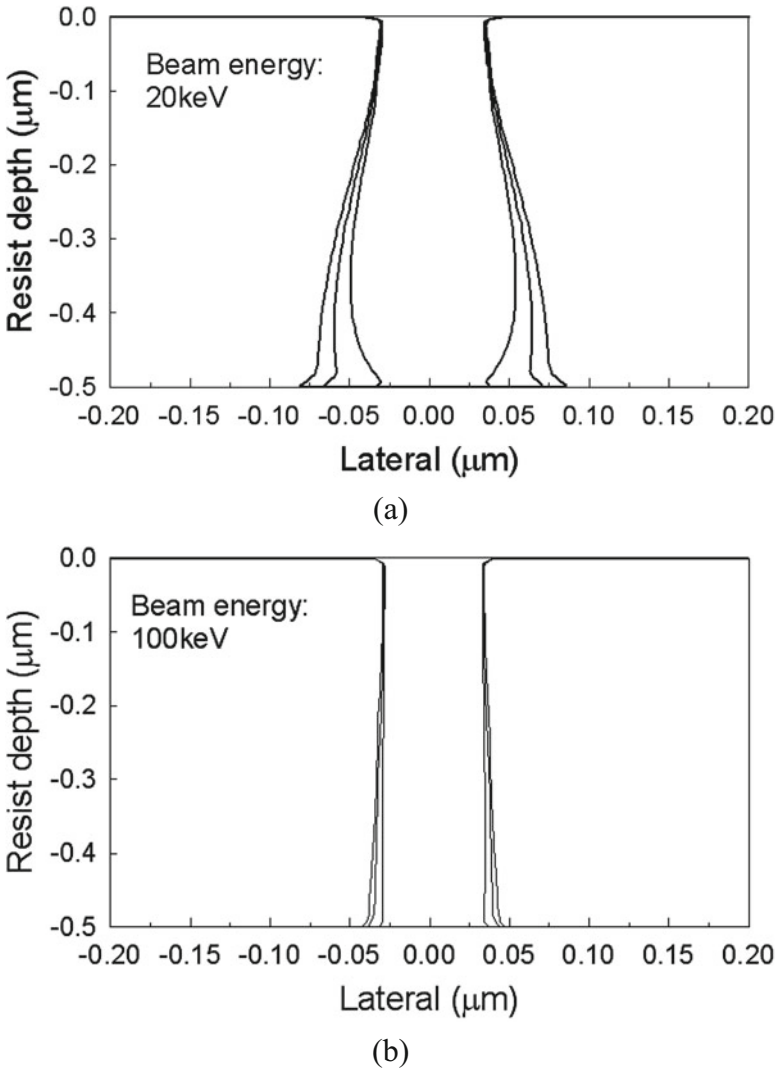


Fig. 8.8 Computer simulated undercut profiles by electron beam exposure: (**a**) single layer PMMA at 20 keV beam energy, three profiles at exposure dose of 100 μC/cm^2, 200 μC/cm^2, and 300 μC/cm^2; (**b**) single layer PMMA at 100 keV beam energy, three profiles at exposure dose of 300 μC/cm^2, 400 μC/cm^2, and 500 μC/cm^2

The bottom layer does not have to be an e-beam sensitive polymer. As long as the top layer is sensitive to e-beam, resist patterning can be performed only on the top layer. Once the top layer is exposed and developed, the bottom polymer can be dissolved by a different solvent through the opening of top layer. By controlling the extent of dissolution for the bottom layer polymer, the length of undercut can be independently controlled. The process is illustrated by Fig. 8.10. One of such

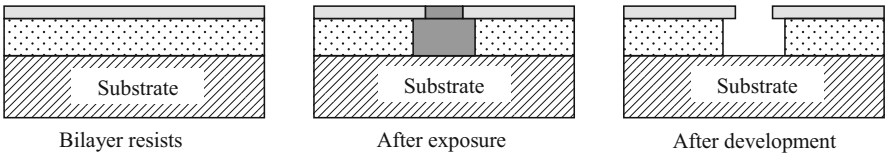


Fig. 8.9 Bilayer resists to create undercut profile for liftoff

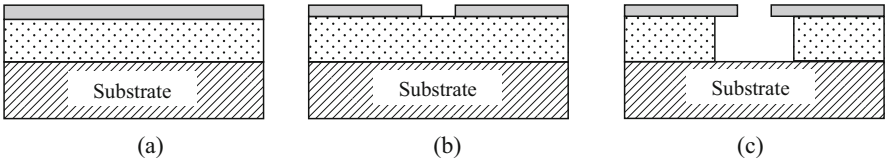


Fig. 8.10 Schematic of bilayer resist system with independent control of undercut: (**a**) bilayer resist stack; (**b**) exposure and development of top layer (imaging layer); (**c**) independent development of bottom layer to control the undercut length

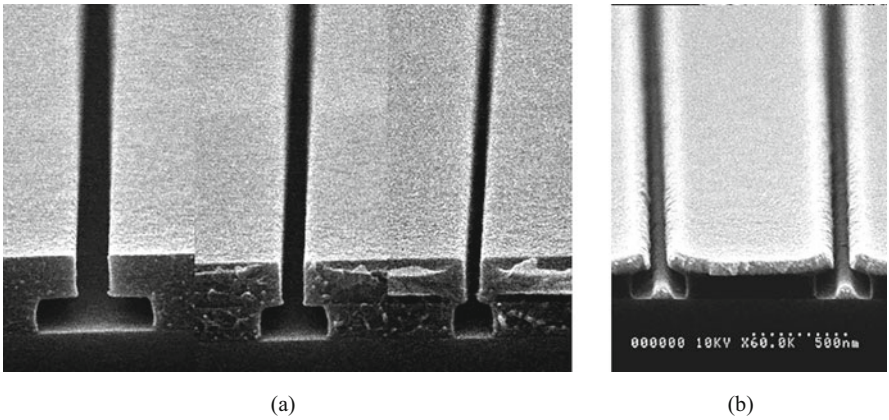


Fig. 8.11 (**a**) Undercut profiles in PMMA/LOR bilayer resist stack; (**b**) metal film deposited on the patterned bilayer resist stack before liftoff

polymers especially designed for liftoff process is the liftoff resist (LOR) based on polydimethylglutarimide (PMGI) polymers which is a standard product available from many resist vendors. Using LOR as the bottom layer in a bilayer resist stack can produce precisely controlled undercut profile as seen in Fig. 8.11a, where three undercut profiles are shown. The length of undercut is controlled only by dissolution time. Figure 8.11b shows deposited metal thin film over the patterned bilayer resist stack before liftoff [10]. With the proper undercut profile, the metal film on top of the resist stack and at the substrate surface is well disconnected.

The same concept can be applied to various other combinations of resists or materials to allow undercut to be independently created. For example, a PMMA/

aluminum bilayer for metal liftoff patterning has been demonstrated with PMMA on top as an imaging layer [11]. The Al underlayer works in the same way as the aforementioned LOR. Al can be dissolved in an alkaline solution, which does not attack the PMMA pattern. The added benefit of using Al underlayer is the ability to carry out e-beam lithography on an insulating substrate, because the aluminum coating can effectively prevent electron charging of the insulating substrate. Silver pattern features with minimum 40 nm linewidth have been produced on a quartz substrate by this process.

It is always difficult to have a completely clean liftoff of metal film, particularly when the pattern area is small and most of the resist layers are covered by metal film. The solvent has to penetrate the metal layer around the gaps at resist pattern edges and dissolve the resist underneath. Once the resist is dissolved, the bulk of metal film must be cleaned away from the sample surface with the assistance of a pressured spray of solvent. The process is slow, and sometimes the metal film can redeposit onto the sample surface once the resist layer is gone. If the metal film is deposited by plasma sputtering which is known to have poor directionality, the film may hang over the pattern edges here and there, resulting in many defects on the sample surface after liftoff. A technique which used a stream of solid CO_2 particles (dry ice), or called carbon dioxide snow, was developed to remove the metal film on top of resist layer prior to a solvent stripping of the resist [12]. The mechanism is to create a thermal expansion mismatch between the metal film and the resist layer because of the cold solid CO_2 below -60 °C. This can cause the metal film to crack and delaminate. The loosened metal film is then cleared away by a CO_2 snow jet with high speed up to 40 m/s, and the resist layer is stripped off in a solvent. The sample backside is usually heated to +60 °C during the process to promote the thermal mismatch and delamination of metal film. Because the metal film is cleared away before the resist layer is dissolved in solvent, the sample surface is well protected by the resist layer, which can prevent any metal film patches from redepositing onto the sample surface.

Most of nanoscale metal pattern liftoff was done by electron beam lithography of positive resist, notably the PMMA resist of either single layer or multilayer. However, there are applications requiring negative tone liftoff, whereby metal thin film is required to cover majority of a substrate surface. In this case, negative tone resist process is preferred because liftoff by positive tone resist such as PMMA will require exposure of large areas and consume long e-beam lithography time. However, negative resist is known to produce overcut profiles, which makes liftoff very difficult. Although some negative photoresists have been used for negative tone liftoff process and undercut profiles can be produced by overdevelopment of the resist structure [13], the process is, however, only suitable for much larger pattern dimension. For high resolution negative tone liftoff, the bilayer concept is still applicable. A HSQ/PMMA bilayer resist process has been developed [14], as illustrated in Fig. 8.12.

HSQ is spin-coated on top of PMMA, and electron beam lithography is performed on the HSQ layer. After development, HSQ negative tone resist patterns are formed. The PMMA will not be affected during the development of HSQ top

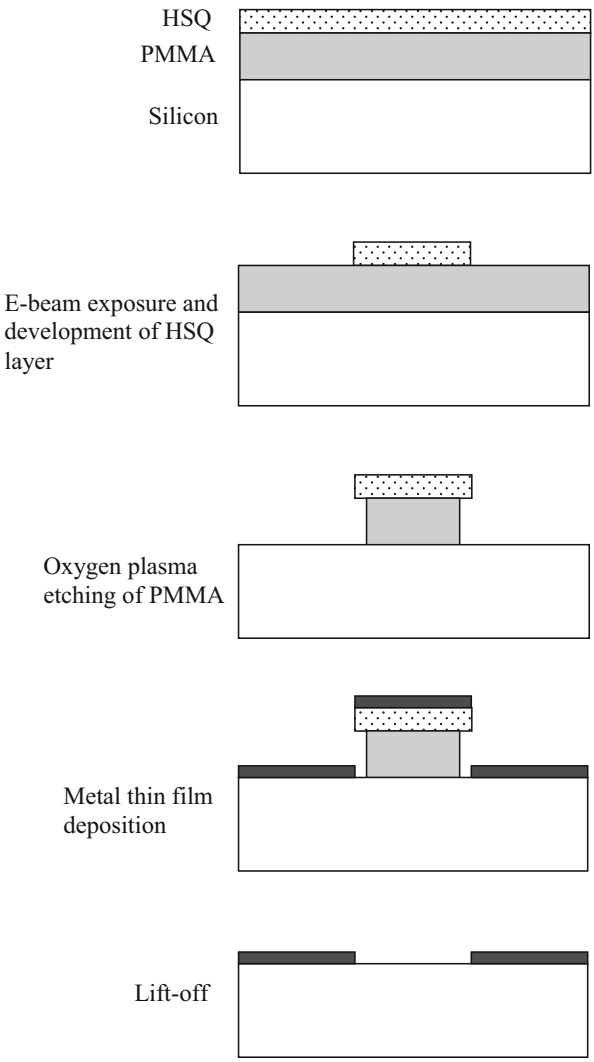


Fig. 8.12 Schematic of negative tone liftoff process using HSQ/PMMA bilayer resist

layer because of the use of incompatible developer, even though it has been exposed by e-beam irradiation through the HSQ layer. The HSQ patterns are then used as an etching mask to transfer into the PMMA underlayer. By controlling the etching parameters, undercut profile of the underlayer PMMA resist can be controlled independently. Figure 8.13 shows the SEM image of HSQ/PMMA resist profile after etching of PMMA underlayer. The negative tone undercut profile is clearly demonstrated. Metal liftoff can be easily done with resist pattern structure. Figure 8.14a shows chrome metal pattern in which each metal line is separated by 50 nm gap which is formed by liftoff of HSQ/PMMA resist lines. Figure 8.14b

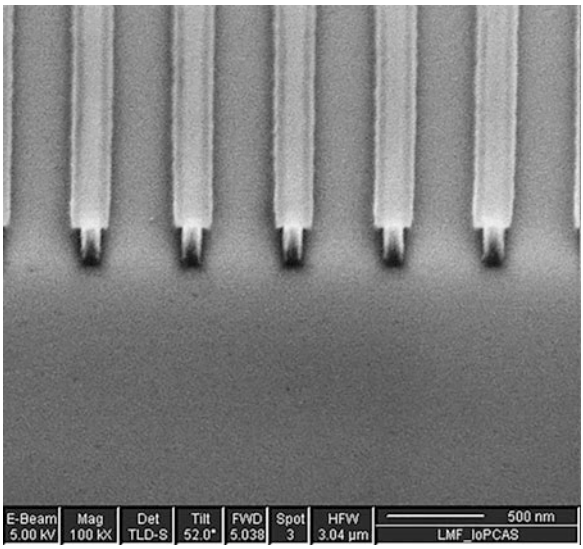
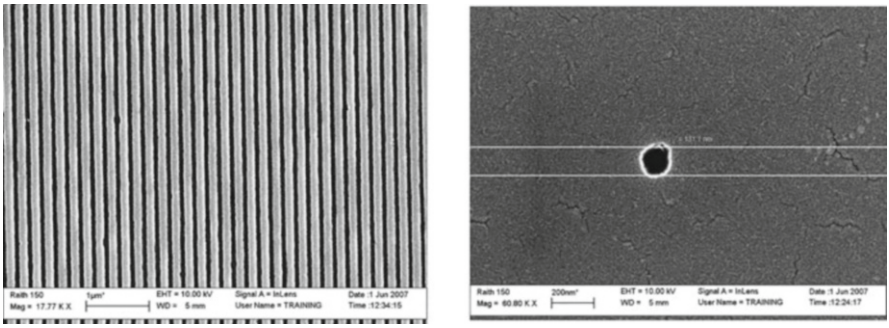


Fig. 8.13 Undercut profile formed from HSQ/PMMA bilayer resist patterning



(a) (b)

Fig. 8.14 Negative tone liftoff patterns: (**a**) 50 nm gaps between chrome; (**b**) 100 nm single hole in chrome film

shows a single 100-nm-diameter hole in a chrome metal film formed from liftoff of HSQ/PMMA resist dot [14].

8.4 Pattern Transfer by Plating

Despite its many advantages in creating metallic patterns, liftoff technique is not suitable to produce high aspect ratio pattern structures. One reason is that the deposited film thickness has to be at most less than one-third of resist depth in order to ensure discontinuity of film, hence a successful liftoff. The other reason is

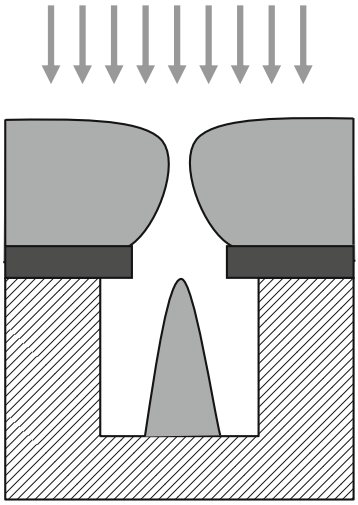


Fig. 8.15 Overhang of deposited materials causing close-up of resist pattern opening

that gradual accumulation of deposited material overhanging at top of resist pattern edge can close up the opening if the resist structure is too narrow and too deep, as illustrated by Fig. 8.15. High aspect ratio metallic micro- or nanostructures are commonly produced by plating.

Plating, either electroplating or electroless plating, is a well-established industrial manufacturing process. Building microstructures by plating started in the 1970s when magnetic storage devices were made by plating copper and permalloy through molds [15]. The molds were photolithographicly defined resist patterns. For electroplating, a seed layer is needed, and electroplating of metallic patterns is done by patterning a resist layer on top of the seed layer and then filling up the patterned resist mold with metal, as illustrated in Fig. 8.1b. Electroless plating is mainly used for conformable coating of a surface. Only a limited number of metals and alloys can be deposited by electroless plating, and the deposition rate is much lower than electroplating [16]. In contrast, many metals and alloys can be electroplated. The major applications of electroplating in microstructure patterning are in thin film head manufacturing for magnetic storage devices [17] and copper filling of via hole for integrated circuits [18]. Many other areas are also using electroplating as an effective method to build high aspect ratio metallic microstructures. Notably, electroplating is one of the main fabrication techniques to make microelectromechanical systems (MEMS).

Electroplating, or electroforming, is one of the key components in the LIGA technique which is a German acronym of "Lithographie (LI) Galvanoformung (G) Abformung (A)," meaning lithography + electroforming + molding [19]. By hard X-ray exposure of very thick polymer resist, microstructures with high aspect ratio have been made by electroplating nickel into the resist mold [16]. Thick photoresist can also be patterned by UV light to form the mold. Though the aspect ratio is not as high as those made by X-ray LIGA, the UV based process is a low cost

version of LIGA and particularly useful for making micromechanical structures [20]. Another main application area where electroplating plays a major role is making various magnetic microdevices by electroplating of magnetic materials, notably the permalloy. Microstructures made of permalloy can be actuated and manipulated by external magnetic fields, for example, magnetic micromirrors which can be deflected by applying a magnetic field externally [21]. The magnetic force exerting on the magnetic microstructures is proportional to the volume of magnetic material. Electroplating is the effective way to build up bulk magnetic materials.

The aforementioned applications using electroplating are mainly for making microscale structures with dimension larger than micrometer or a few hundreds of micrometers. Electroplating for making sub-100 nm structures is primarily in two areas: first area is in making masks and zone plate lens for X-ray lithography or X-ray imaging; second area is in making stamps for nanoimprinting, which has been described in Sect. 6.2.1 of Chap. 6. For X-ray lithography, heavy metal structures act as absorbers in an X-ray mask. They have to be thick enough to absorb all the X-ray radiation, as have been described in Sect. 2.4.3 of Chap. 2. As X-ray is 1:1 proximity imaging, to pattern sub-100 nm structures by X-ray lithography, the absorber pattern on the mask has to be sub-100 nm as well. Micro zone plate (MZP) is a common type of imaging lens to focus X-ray radiation. The resolution of a MZP lens can be expressed as Eq. 8.5 [22], which is similar to that of optical imaging (Eq. 2.3 in Chap. 2):

$$R_x = k_1 \frac{\lambda}{\mathrm{NA_{MZP}}}, \quad \text{where} \quad \mathrm{NA_{MZP}} = \frac{\lambda}{2\Delta r_{\mathrm{MZP}}} \tag{8.5}$$

where λ is the X-ray wavelength, $\mathrm{NA_{MZP}}$ is the numerical aperture of the zone lens, and $\Delta r_{\mathrm{MZP}}$is the outmost zone width. k_1 is illumination-dependent constant. Then the resolution can be expressed as $2k_1\,\Delta r_{\mathrm{MZP}}$. With an optimized X-ray illumination system, the k_1 factor can be less than 0.5 [23]. It is apparent that the imaging resolution is solely dependent on the outmost zone width. The smaller the zone width, the higher the imaging resolution. Making high resolution MZP lens is a great challenge, and electroplating is the only method to achieve high aspect ratio (requirement of opacity to X-ray) heavy metal structure.

The material for high aspect ratio electroplating at sub-100 nm dimension is mainly gold. Electroplated gold is in the form of polycrystalline structure which can have many grain boundaries and imperfections. The polycrystalline grains may grow faster in a particular orientation. For example, gold plating is faster in the <111> direction [24]. The grains usually grow perpendicular to the surface. Therefore, surface roughness is the common feature of electroplated patterns. Figure 8.16 shows a part of MZP lens which is made from electroplated gold. The linewidth of outmost ring is 25 nm. Polycrystalline grains are visible from the surface of MZP. If the resist pattern dimension is comparable to the grain size, voids or even break can happen to the electroplated lines if the height of feature is much greater than its width [24]. Figure 8.17 shows a MZP with outmost nominal line width of 15 nm. As seen

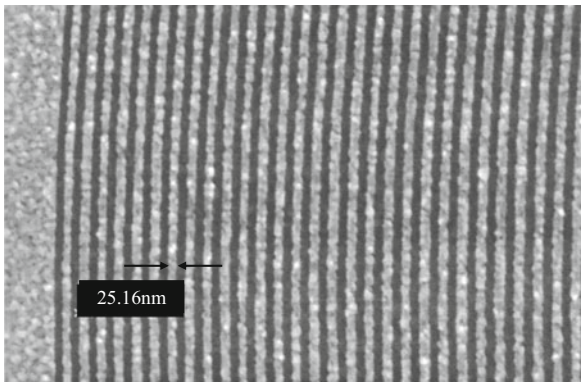


Fig. 8.16 Zone plate with 25 nm electroplated gold lines in the outmost zone. (Reprinted from [25] with permission)

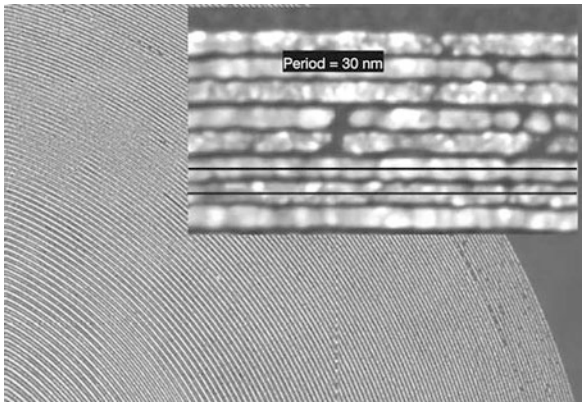


Fig. 8.17 Zone plate with 15 nm electroplated gold lines in the outmost zone (inset image showing detail of 15 nm gold lines). (Reprinted from [23] with permission)

from the inset SEM image, the gold lines are wider than its nominal 15 nm width, which is the result of proximity effect in e-beam lithography. In addition, some of the gold lines are visibly broken, which is most likely due to gain size effect in gold electroplating.

A critical issue in pattern transfer by electroplating is the uniformity of electroplated structures. As the deposition rate in electroplating is highly dependent on plating current, the uniformity of deposition thickness depends on the uniformity of current density distribution. The current density is always higher at the location where there are asperities such as the boundary of a substrate or sharp edge of a pattern structure. The consequence of such nonuniform distribution of current density is the nonuniformity of the thickness of electroplated features. Generally, large electroplated features may be thicker at the edges. Isolated narrow lines are thicker than densely packed narrow lines [16]. For a large substrate with many different sizes of features, it often happens that some features have reached the correct thickness while others may have been over plated.

An important additive pattern transfer process in IC manufacturing is the damascene process. The name *damascene* came from its similarity to ancient jewelry making process which inlays metal wires and patterns. IC manufacturing involves multilevel interconnecting transistors to make a circuit functioning. Interconnects were done by depositing a metal film (Al or Cu) and patterning the metal film into connecting lines. Between different levels of interconnects, via holes are etched and filled with metal to connect the top and bottom levels of metal lines. In original single-damascene process, via hole filling and metal line patterning are done separately. The single-damascene process is suitable for Al interconnects because Al can be easily etched by RIE. With interconnects moving from Al to Cu, the poor dry etchability of Cu made the single-damascene process unattainable. The dual-damascene process was then developed, which uses a single metal deposition process to simultaneously form the metal lines and metal vias [26].

The schematic cross section of copper interconnects and vias is shown in Fig. 8.18a [27]. It starts first with nanoscale vias and trenches etched in an interlayer dielectric (typically SiO_2) and followed by deposition of diffusion barrier layers. A thin Cu seed layer is then deposited by plasma sputtering, followed by Cu electroplating. Finally, excessively plated Cu in the field regions is removed by chemical mechanical polishing (CMP). Figure 8.18b is the cross-sectional SEM image of interconnected IC showing the filled oxide dielectrics and the copper vias as well as interconnects. In recent years, ICs are developing toward 3D stack and system on chip (SoC) which need through-silicon-via (TSV) to interconnect different levels of circuits. The TSVs normally have depth from 10 μm to more than 100 μm and aspect ratio >20:1. Fig. 7.20b in Chap. 7 has shown such a TSV which is 95 μm deep and filled with tungsten. At such high aspect ratio, the TSVs have to be etched by DRIE process, and the seed layer has to be deposited by ALD. Copper or tungsten then fills the TSV by electroplating or CVD process [28].

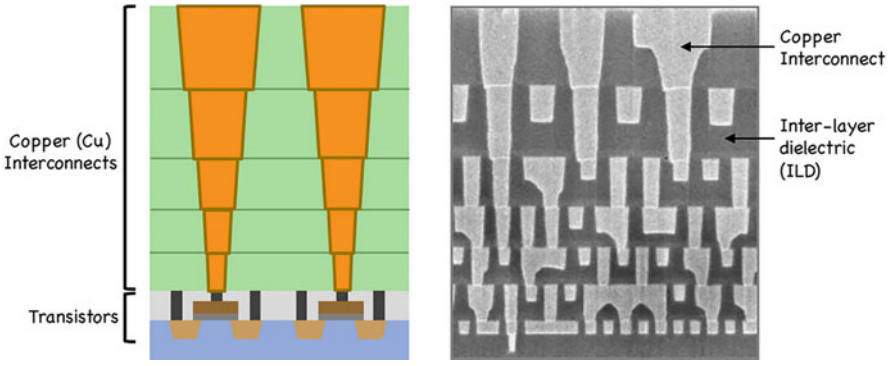


Fig. 8.18 (**a**) Schematic cross section of integrated circuit with vias and interconnects. (**b**) Image of copper interconnects in an integrated circuit. (Reprinted from [27] with permission)

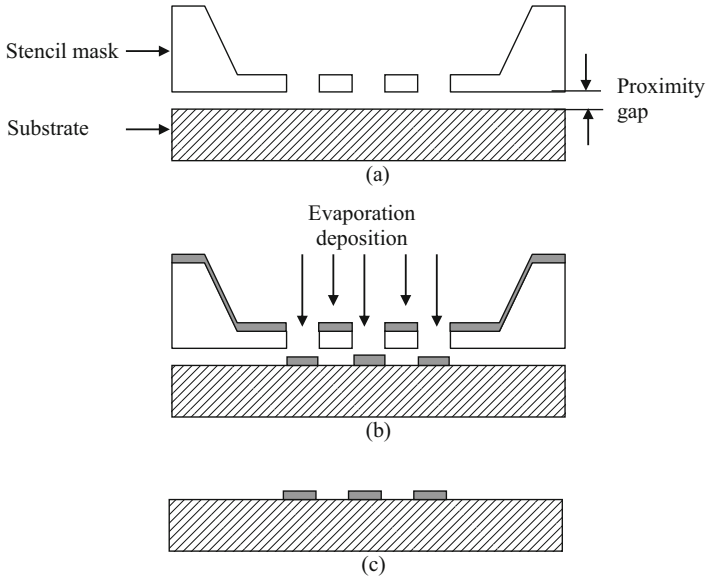


Fig. 8.19 (**a**) Stencil mask and substrate configuration; (**b**) thermal evaporation deposition through stencil mask; (**c**) pattern transfer to substrate

8.5 Pattern Transfer by Stencil Mask

Metallic patterns can also be transferred directly onto a substrate by a stencil mask. The process is illustrated in Fig. 8.19. A stencil mask is a framed membrane with etch-through patterns. Pattern transfer by stencil mask is to use a stencil mask to partially block material deposition on a substrate. Deposited material forms a pattern on a substrate, which is the same as the etch-through pattern on the stencil mask. Or rather, it leaves a shadow of the mask on substrate during material deposition. So it can also be called shadow mask patterning. To ensure the shadowing effect, the deposition has to be directional. Therefore, thermal evaporation deposition is the only viable deposition technique to combine with the stencil mask pattern transfer process.

Nanoscale pattern transfer through stencil mask was first reported in 1999 [29]. The stencil mask was made of a suspended silicon nitride membrane with holes of 5–10 nm. Thermally evaporated metal dots of 15 nm were made through the membrane mask. Later large area patterning was developed [30, 31]. Stencil mask patterning has the following advantages.

- A thin film pattern can be deposited at any local position, because the stencil mask is not in contact with substrate. Pattern deposition can also be done on topographical surface.

- Thin film patterns are formed in one step, without resist patterning and liftoff processes like conventional lithography. Because no resist patterning is needed, material can be deposited directly on fresh substrate surface, avoiding potential surface contamination. As the deposition is carried out in vacuum, the possibility of substrate surface oxidation can also be excluded. Mechanically fragile or chemically functionalized surface can be structured due to the noncontact and process-free nature of the pattern transfer.
- Because no optical or e-beam lithography is involved, there is no proximity effect caused either by light diffraction or electron scattering. Large and small patterns can be formed simultaneously. For example, it has been reported that on the same stencil mask, both sub-100 nm features and above 100 μm features were made in one step deposition [31].

However, when transferring patterns by stencil masks, some cautions have to be taken as there are limitations in this technique.

- The technique can only deposit single layer pattern. For multilayer deposition, there is an alignment issue, unless the multilayer materials are deposited on the same location.
- For thick film deposition or repeated use of stencil mask, the deposition of material around the edge of stencil pattern can close in, resulting in smaller feature dimension on the mask, and eventually clog some small holes completely. Figure 8.20a shows schematically the ideal deposition without clogging, and Fig. 8.20b is the deposition with clogging at the stencil mask aperture. The eventual close-up of mask aperture is demonstrated by the SEM image of real stencil mask in Fig. 8.20c. Nevertheless, it is possible to remove clogging by wet chemical or reaction ion etching after each time use. It is the clogging during the deposition that imposes a limit on the amount of material to be deposited in a single process step.
- Because there is a gap between stencil mask and substrate surface, there are always blurs like halos around the main pattern edge, due to molecules arriving in angular direction through the stencil mask opening. Further blurs are caused by kinetic movement of molecules, because the molecules that evaporated thermally from the material source have kinetic energy and will move laterally before landing on the substrate surface. The blur, however, is much thinner than the main deposited pattern structure and can be corrected by an extra etching after the stencil mask pattern transfer [32].
- For large size stencil mask, there is a thermal stability issue. Deposition of metal by thermal evaporation can generate a lot of heat from evaporation source. Radiation of heat can cause expansion of stencil mask, causing pattern shift and distortion. The deposition induced stress can also deform the stencil membrane.
- The fabrication of stencil mask involves other lithographic techniques and reactive ion etching, which limits the smallest feature achievable on the mask. For large area stencil mask, the minimum mask features are around 100 nm. For small size stencil mask, focused ion beam sputtering can be used to make very small holes in membrane.

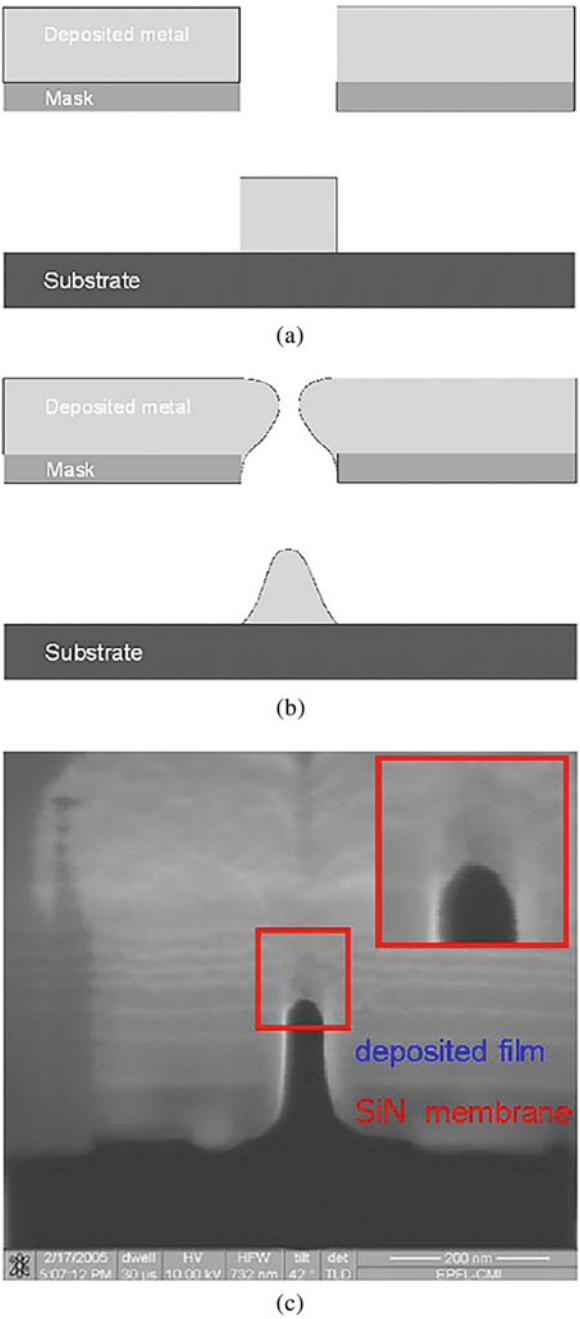


Fig. 8.20 (**a**) Ideal deposition through stencil mask aperture without clogging; (**b**) deposition with clogging; (**c**) SEM image of a stencil mask aperture which is closed up due to excess deposition. (Reprinted from [33] with permission)

8.6 Pattern Formation by Printing

The additive pattern transfer technologies described in the previous sections are all based on transfer through a mask, either a polymer resist mask or a stencil mask. Functional materials can also be deposited into patterns by printing, such as inkjet printing, screen printing, and gravure printing. Patterning a polymer resist mask needs to go through a lengthy lithography process, including spin-coating resist, lithographic patterning (optical or e-beam), developing the exposed resist, etc. Making a stencil mask is even more complicated, which needs not only resist patterning but also metal etching. Printing, especially inkjet printing, does not need a mask but deposits functional materials directly into designed patterns. Additive pattern formation by printing has a number of advantages over lithography-based technology:

- Printing can deposit materials on much larger substrate than conventional lithography which is so far limited to 300 mm size of a silicon wafer.
- Printing has no restriction on the type of substrate to deposit materials, particularly on various flexible substrates which conventional lithography cannot accommodate.
- Printing is a low cost material patterning process, not only because the printing equipment is low cost but also because of its high throughput capability which greatly reduces the unit cost of large volume production.
- For inkjet printing, there are added benefits that the pattern formation is entirely digital without the need for a mask, and it is on-demand deposition of materials with little wastes.

Figure 8.21 schematically compares inkjet printing with shadow mask deposition for manufacturing organic light emitting diode (OLED) display panels. While thermal evaporation of colored pixels (red-blue-green) through a shadow mask wastes 90% of precious organic materials, inkjet printing performs on-demand deposition which saves 90% of the material. Conventionally, all the OLED TV panels are made by shadow mask deposition for the color pixels. In May 2023, TCL, the leading display panel manufacturer in China, unveiled its inkjet-printed 65-inch flexible OLED TV, which demonstrated the capability of printing to replace traditional additive pattern formation technologies. In fact, printing functional materials, such as conductors, semiconductors, or dielectric materials, to fabricate electronic devices and circuits has become a new engineering discipline in recent years, called printed electronics [34].

From the nanofabrication point of view, the pattern formation with traditional printing is not of nanoscale. The existing printing techniques, including inkjet, gravure, gravure offset, or screen printing, print features in micrometers or hundred micrometer dimension. Taking the example of inkjet printing, traditional inkjet printers, as we use in office in everyday, are only capable of printing features no less than 50 μm. It relies on jet out ink droplets from a nozzle by piezoelectric pushing, as schematically shown in Fig. 8.22a. The dimension of printed feature

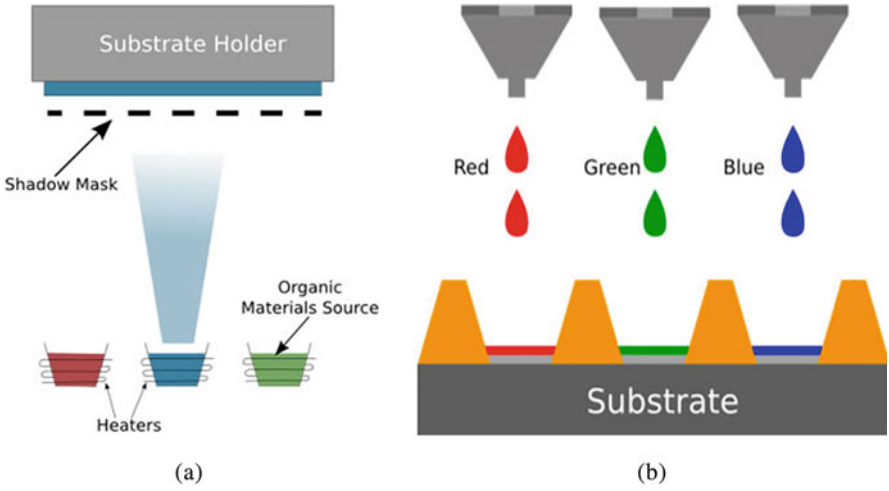


Fig. 8.21 Color pixel formation through (**a**) a shadow mask and (**b**) by inkjet printing

depends largely on the size of ink droplets which in turn depends on the size of nozzle. However, the inkjet nozzle cannot be made very small because of the ink clogging problem. The smallest nozzle that can be safely used is about 20 μm, which determines the low limit of feature dimension that can be printed by this type of inkjet printers. The small driving force by piezoelectric pushing in a traditional inkjet printer also permits only low ink viscosity.

The demands for high resolution printing in printed electronics greatly stimulated the development of new printing methods. An advance in inkjet printing technology is the development of aerosol jet printing [35], as schematically shown in Fig. 8.22b. In an aerosol jet printer, inks are atomized into aerosol which is blown out by high speed airstream. At the exit of the nozzle, a ring of compressed air aerodynamically focuses the aerosol stream into a much finer jet. Although the diameter of nozzle can be 200 μm, the focused jet stream is as small as 5 μm, ensuring much finer printed features. The much larger nozzle used in the aerosol jet printer can also avoid the clogging problem due to high viscosity ink or dispersed nanomaterials such as carbon nanotubes or metal nanowires in an ink. The breakthrough in improving the resolution capability of inkjet printing is the development of electrohydrodynamic jet (E-jet) printing [36]. The principle of E-jet is illustrated in Fig. 8.22c. The ink droplets are pulled out by an electric field between the nozzle and substrate. A fine jet is achieved because a liquid under the influence of electric field can form a cone shape with very sharp end. Further increasing the field can pull out the apex of liquid cone as fine droplets or a stream. With E-jet printing technology, high aspect ratio patterns with 50 nm lateral resolution have been demonstrated.

Printing can be considered as a different kind of nanomanufacturing process, not characterized by its patterning resolution but by its ability to deposit patterned

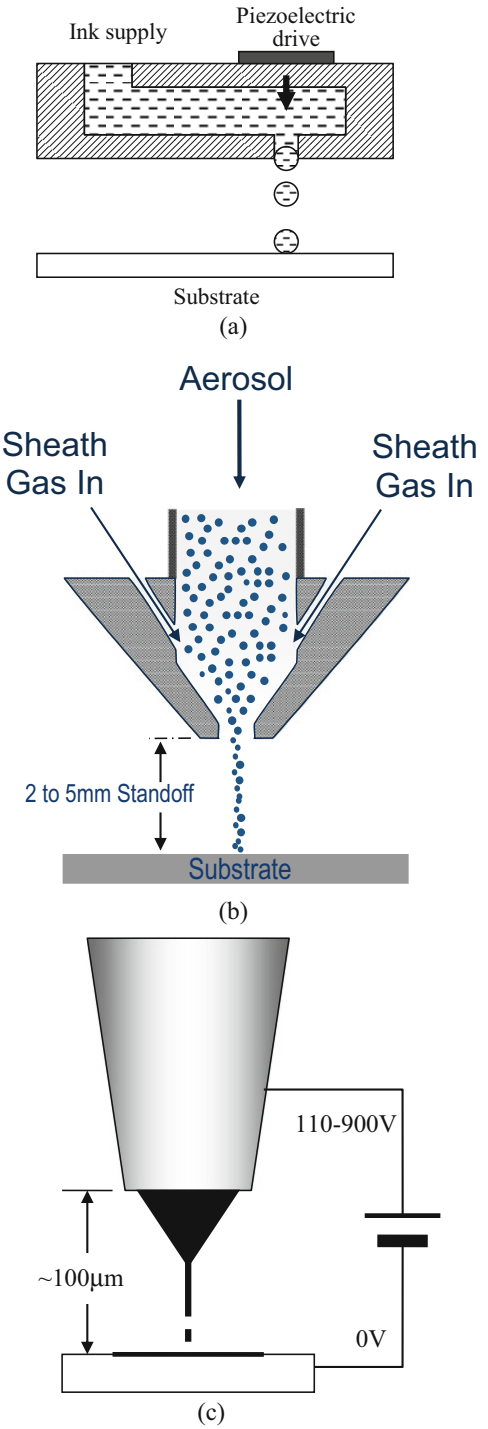


Fig. 8.22 Schematics of different inkjet printing technologies: (**a**) traditional inkjet; (**b**) aerosol jet; (**c**) electrohydrodynamic jet (E-jet)

nanomaterials. Although a wide variety of nanomaterials have been developed over the past decades, including a variety of nanoparticles and nanowires, carbon nanotube, graphene, and a range of 2D nanomaterials, the applications of these nanomaterials in electronic devices have not been very successful. One of the key reasons is the integration issue. The extraordinary functions of nanomaterials only work on individual entity, such as single nanowire, single carbon nanotube, or a single layer graphene. Although excellent working devices have been demonstrated by integrating single entity of nanomaterial into devices in laboratories, they cannot be done in large scale or as mass manufacturing process. Printing offers a solution [37]. By formulating the nanomaterials into inks and printing them, large scale integration of nanomaterials into devices becomes possible. Though the printing does not have fine resolution as conventional nanofabrication, many electronic devices, such as sensors, displays, or solar cells, do not have to be as small as those in an integrated circuit chip. Once the nanomaterials are printed, they become part of electronic devices and play a key role with their nanoscale properties. The key aspect of printing is that large scale manufacturing can be done with high throughput and low cost. It opens a new avenue for widening the application of nanomaterials.

References

1. Elshabini-Riad, A.A.R., Barlow III, F.D.: Thin Film Technology Handbook. McGraw-Hill (1997)
2. Johnson, R.W., et al.: A brief review of atomic layer deposition: from fundamentals to applications. Mater. Today. **17**(5), 236–246 (2014)
3. Chambers, A., Fitch, R.K., Halliday, B.S.: Basic Vacuum Technology, 2nd edn. CRC Press (1998)
4. Chapman, B.: Glow Discharge Processes. John Wiley & Sons, Inc (1980)
5. Rossnagel, S.M.: Sputter deposition for semiconductor manufacturing. IBM J. Res. Develop. **43**(1/2), 163 (1999)
6. Rossnagel, S.M., et al.: Collimated magnetron sputter deposition. J. Vac. Sci. Technol. A. **9**, 261–265 (1991)
7. Ilataakis, M., Canavcuo, Shaw, I.M.: Single-step optical lift-off process. IBM J. Res. Develop. **24**(4), 452–460 (1980)
8. Beaumont, S.P., et al.: Sub-20-nm-wide metal lines by electron-beam exposure of thin poly (methyl methacrylate) films and liftoff. Appl. Phys. Lett. **38**(6), 436–439 (1981)
9. Chou, S.Y., et al.: Sub-10 nm imprint lithography and applications. J. Vac. Sci. Technol. **B15**(6), 2897 (1997)
10. Chen, Y., Peng, K., Cui, Z.: A lift-off process for high resolution patterns using PMMA/LOR resist stack. Microelectron. Eng. **73/74**, 278 (2004)
11. Xia, X., et al.: Fabrication of near-infrared and optical meta-materials on insulating substrates by lift-off using PMMA/Al stack. Microelectron. Eng. **84**(5–8), 1144–1147 (2007)
12. Radulescu, F., et al. Introduction of Complete Sputtering Metallization in Conjunction with CO2 Snow Lift-Off for High Volume GaAs Manufacturing. In: Proc. GaAs MANTECH Conference. 2002
13. Voigt, A., et al.: A single layer negative tone lift-off photo resist for patterning a magnetron sputtered Ti/Pt/Au contact system and for solder bumps. Microelectron. Eng. **78–79**, 503–508 (2005)

14. Yang, H., et al.: Electron beam lithography of HSQ/PMMA bilayer resists for negative tone lift-off process. Microelectron. Eng. **85**(5–6), 814-817 (2008)
15. Romankiw, L.T., Croll, L.M., Hatzakis, M.: Batch fabricated thin film magnetic recording head. IEEE Trans. Mag. **6**(4), 729 (1970)
16. Romankiw, L.T., O'Sullivan, E.J.M.: Plating techniques. In: Rai-choudhury, P. (ed.) Handbook of Micolithography, Micromachining and Microfabrication. SPIE Optical Engineering Press & IEE (1997)
17. Jiang, M.: Processing considerations for CMP on thin-film head wafers. Solid State Technol. **47**, S17–S19 (2004)
18. Jackson, R.L., et al.: Processing and integration of copper interconnects. Solid State Technol, 49–59 (1998 March)
19. Becker, E.W., et al.: Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming and plastic moudling (LIGA process). Microelectron. Eng. **4**, 35 (1986)
20. Cui, Z., Lawes, R.A.: Low cost fabrication of micromechanical systems. Microelectron. Eng. **35**, 389 (1997)
21. Cui, Z., et al.: High sensitive magnetically actuated micromirrors for magnetic field measurement. Sensors Actuators A Phys. **138**(1), 145–150 (2007)
22. Attwood, D.T.: Soft X-rays and extreme ultraviolet radiation. Cambridge University Press (2000)
23. Chao, W., et al.: Soft X-ray microscopy at a spatial resolution better than 15 nm. Nature. **435**, 1210 (2005)
24. Simon, G., et al.: Electroplating: an alternative transfer technology in the 20 nm range. Microelectron. Eng. **35**, 51–54 (1997)
25. Anderson, E.H., et al.: Nanofabrication and diffractive optics for high-resolution x-ray applications. J. Vac. Sci. Technol. **B18**, 2970 (2000)
26. Wolf, S.: Introduction to dual-damascene interconnect processes. Silicon Process. VLSI Era. **4**, 674–679 (2004)
27. Akolkar, R.: Current status and advances in damascene electrodeposition. In: Wandelt, K. (ed.) Encyclopedia of Interfacial Chemistry: Surface Science and Electrochemistry, pp. 24–31. Elsevier (2018)
28. Gambino, J.P., et al.: An overview of through-silicon-via technology and manufacturing challenges. Microelectron. Eng. **135**, 73–106 (2015)
29. Deshmukh, M.M., et al.: Nanofabrication using a stencil mask. Appl. Phys. Lett. **75**(11), 1631–1633 (1999)
30. Brugger, J., et al.: Resistless pattering of sub-micron structures by evaporation through nanostencils. Microelectron. Eng. **53**, 403 (2000)
31. Kim, G.M., Van den Boogaart, M.A.F., Brugger, J.: Fabrication and application of a full wafer size micro/nanostencil for multiple length scale surface patterning. Microelectron. Eng. **67–68**, 609 (2003)
32. Arcamone, J., et al.: Dry etching for the correction of gap-induced blurring and improved pattern resolution in nanostencil lithography. J. Micro/Nanolithogr. MEMS MOEMS. **6**(1), 013005 (2007)
33. Lishchynska, M., et al.: Predicting mask distortion, clogging and pattern transfer for stencil lithography. Microelectron. Eng. **84**, 42–53 (2007)
34. Cui, Z.: Printed Electronics: Materials, Technologies and Applications. Wiley (2016)
35. Hines, D.R., et al.: Considerations of aerosol-jet printing for the fabrication of printed hybrid electronic circuits. Addit. Manuf. **47**, 102325 (2021)
36. Mkhize, N., Bhaskaran, H.: Electrohydrodynamic jet printing: introductory concepts and considerations. Small Sci. **2**, 2100073 (2022)
37. Gonzalez, V., Puente, C., Lopez, I.: Nanomaterials for printing technology: a revolution in industrial manufacturing. In: Hussain, C.M. (ed.) Handbook of Nanomaterials for Industrial Applications, pp. 1012–1030. Elsevier Inc. (2018)

Chapter 9
Indirect Nanofabrication



9.1 Introduction

The nanofabrication technologies introduced in the previous chapters so far have a common feature, that is, the dimension of a structure or pattern is defined directly by a lithographic means, whether through direct patterning (electron beam, focused ion beam, laser beam, scanning probes, inkjet printing), or through a mask (photomask or stencil mask) or by a mold (nanoimprint or soft lithography), and/or combined with pattern transfer such as etching or liftoff. Though each lithographic technique has its own capability and limit, most of aforementioned techniques are capable of sub-10 nm patterning resolution after years of development both in lithography equipment and in patterning materials as well as processes [1].

It should be aware that to achieve such ultimate resolution capability, one is assumed to have access to the expensive state-of-the-art equipment, which may be a luxury that many do not have. Some nanofabrication practitioners may have to go with whatever equipment they can access and try their best to explore the potential of their existing fabrication facility or to find ways to get round of the resolution limit of their own equipment. The indirect nanofabrication technologies to be introduced in this chapter are those clever ways of getting round of the limits of existing fabrication equipment. They are indirect because the pattern dimension is not predefined by a direct lithographic technique. Here is a simple example; if one cannot make a very narrow line pattern, say sub-10 nm linewidth, by optical or e-beam or scanning probe, one can start with a wide line pattern and take away bit by bit from both sides until it becomes very narrow. This is the sort of indirect nanofabrication technique to be discussed, and the principle can be applied to many other ways of narrowing down a pattern structural dimension.

It is difficult to categorize each indirect nanofabrication techniques because there can be many diverse methods to overcome the existing fabrication limits. The above example can be considered as an extension of etching techniques introduced in Chap. 7. It is a lateral etching instead of vertical etching, without direct involvement

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_9

of lithographic patterning. There are also methods utilizing material deposition to create pattern structures whose lateral dimension is not defined by vertical lithography but by the thickness of deposited film on the sidewall of a patterned structure. Generally, indirect nanofabrication is not a straightforward "lithography plus pattern transfer" process but rather more complex, and several process steps may be involved. In most cases, they do not have the flexibility and versatility enjoyed by direct nanofabrication technologies, for example, making arbitrary shapes of patterns can be difficult with indirect techniques. Nevertheless, indirect nanofabrication is a research area where many tricks have been played to push down the limit of existing nanofabrication capabilities.

9.2 Sidewall Lithography

Sidewall lithography was the name given to a technique which achieves narrow width of pattern structure by thin film coating on the sidewalls of supporting structures and selectively etching away the supporting structures. The technique is schematically explained in Fig. 9.1. The supporting pattern structures are first made with low-resolution direct lithographic process, for example, optical lithography (Fig. 9.1a). The second step is to conformably coat a thin film covering the supporting structure (Fig. 9.1b). The thin film has to be a material different from the supporting structure. A selective RIE process removes the top thin film cover of the supporting structure (Fig. 9.1c). The final step is to selectively etch away the supporting structure, leaving the sidewall thin film standing alone as the final pattern (Fig. 9.1d). Figure 9.2 shows the SiO_2 structures created by conformal coating SiO_2 onto silicon supporting structure and selective etching [2]. It is apparent that the lateral width of final pattern structure depends only on the thickness of thin film coated on the sidewalls of supporting structures. So is the name of sidewall lithography. There are two critical requirements that enable the sidewall lithography:

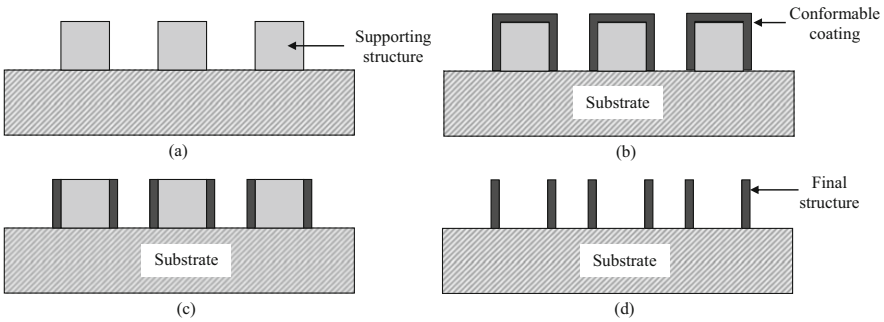


Fig. 9.1 Schematic of sidewall lithography process: (**a**) patterning supporting structure; (**b**) conformable coating; (**c**) removal of top cover thin film by RIE; (**d**) removal of supporting structure

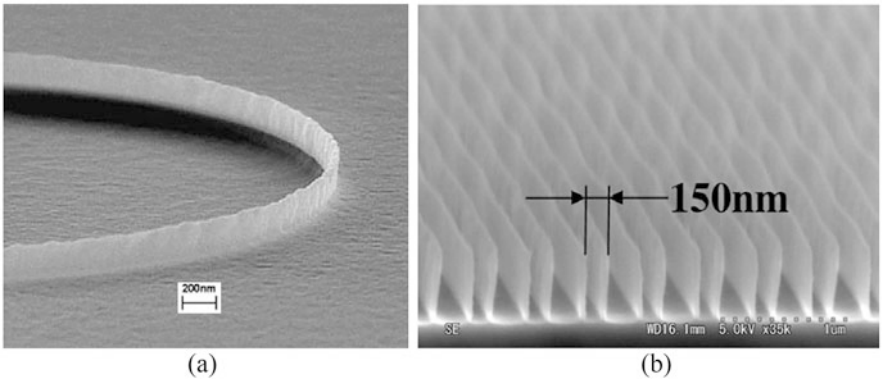


Fig. 9.2 (**a**) Circular SiO_2 structure and (**b**) SiO_2 grating structure made from removing silicon supporting structures. (Reprinted from [2] with permission)

(1) the supporting structures must have a vertical sidewall, so that when the top cover is removed by RIE, there is a minimum removal of the sidewall coating material; (2) the two materials for the thin film and the supporting structure have to be incompatible in RIE or chemical wet etching, so that the supporting structure can be removed without damaging the sidewall thin film structure.

Fine resolution structures created by sidewall deposition were first reported in 1983, which was then called edge defined technique [3]. It consisted of repeatedly applied CVD deposition, reactive ion etching, and selective isotropic etching with a proper choice of materials and etching chemistry. Each cycle doubles the number of lines in the pattern. Gratings with ~40 nm period was then fabricated by doubling the number of lines in a 160 nm period grating twice. Later a much simple process was proposed, in which the supporting photoresist was silylated to replace the conformable thin film coating step [4]. A minimum of 50 nm wide and 200 nm high silicon nitride patterns had been fabricated by RIE transfer of sidewall patterns, which was inconceivable at the time by any other lithographic techniques.

Since it was reported, sidewall lithography has been continuously employed till present day to make fine pattern structures without using high resolution lithography tools [5, 6]. The basic concept is the same. The only variations are different combinations of thin film and supporting structure materials. There have been reports using platinum coating on amorphous silicon to make 50 nm Pt patterns [7], using SiO_2 as the supporter and Si_3N_4 as the thin film to make 100 nm line and space grating for nanoimprint [8], using polysilicon and SiO_2 combination to make sub-10 nm SiO_2 pattern structures as masks to etch into silicon [9], or using patterned silicon as supporting structures and thermal oxidation process to modify the sidewall surface of the silicon supporters while protected by Si_3N_4 at the top [10]. Whatever the combination of materials is, as long as the two aforementioned critical requirements are observed, sub-100 nm patterns can be made with low cost, albeit a slight increase in process complexity. However, the flexibility of making any shapes of 2D

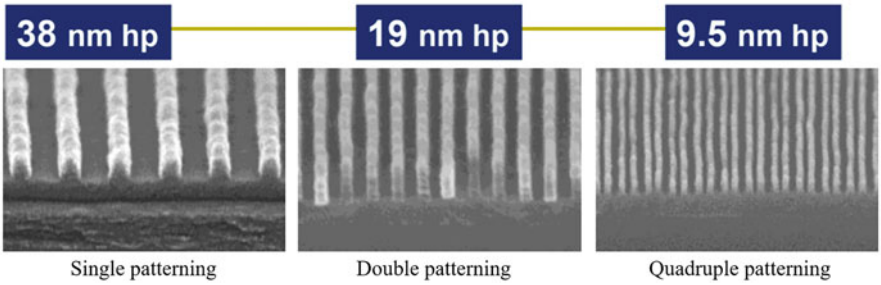


Fig. 9.3 Multiple patterning by self-aligned spacer process. (Reprinted from [12] with permission)

pattern layout is restricted with the technique, because the final pattern layout relies on the pattern of supporting structures as shown in Fig. 9.2.

The most prominent example of sidewall lithography is the application to IC manufacturing. As described in 2.7.5 of Chap. 2, deep UV optical lithography with 193 nm wavelength is only able to pattern at 130 nm resolution. With immersion lens technique, the resolution can be pushed down to 38 nm. Further down scaling from 14 nm node IC manufacturing onward, multiple patterning processes were used when EUV lithography was not ready. One of the dominant multiple patterning processes is self-aligned spacer technology which is exactly the same as the sidewall lithography illustrated in Fig. 9.1, if comparing it with the self-aligned double patterning (SADP) process in Fig. 2.40. With the help of spacer which is the sidewall deposition, pattern density can be doubled [11]. Figure 9.3 is the demonstration that 38 nm half-pitch (hp) line pattern by single 193 nm immersion lithography was reduced to 19 nm hp. by self-aligned double patterning (SADP) and further reduced to 9.5 nm by self-aligned quadruple patterning (SAQP) [12]. By self-aligned octuplet patterning (SAOP), the half-pitch was eventually reduced to 5.5 nm [13].

While the sidewall lithography relies on selective removal of supporting structures to leave the sidewall thin film as the final pattern structures, the supporting structures can be left where they are, and the sidewall thin film serves as a nanoscale insulator. Figure 9.4 illustrates the concept of adhesion lithography which creates nanoscale insulating gaps between metal electrodes [14]. The process starts with depositing the supporting structure which is a metal (M1) as indicated in Fig. 9.4a. A metallophilic self-assembled monolayer (SAM) is then deposited to cover the whole substrate including the M1 structure (Fig. 9.4b). A second metal (M2) is deposited to cover the whole substrate again (Fig. 9.4c). Now an adhesive tape is applied to the whole substrate surface (Fig. 9.4d), and then part of M2 which is in contact with the SAM region is peeled away due to the weak adhesion with M1 because of the SAM coating (Fig. 9.4e). Finally, the substrate is left with two metallic structures (M1, M2), as seen in Fig. 9.4f where the SAM layer in between serves as an insulator which is only 12 nm wide averagely. Such nanoscale gaps are created without costly equipment and under ambient conditions on a timescale of seconds. The insulating layer can also be deposited by atomic layer deposition (ALD) on a patterned

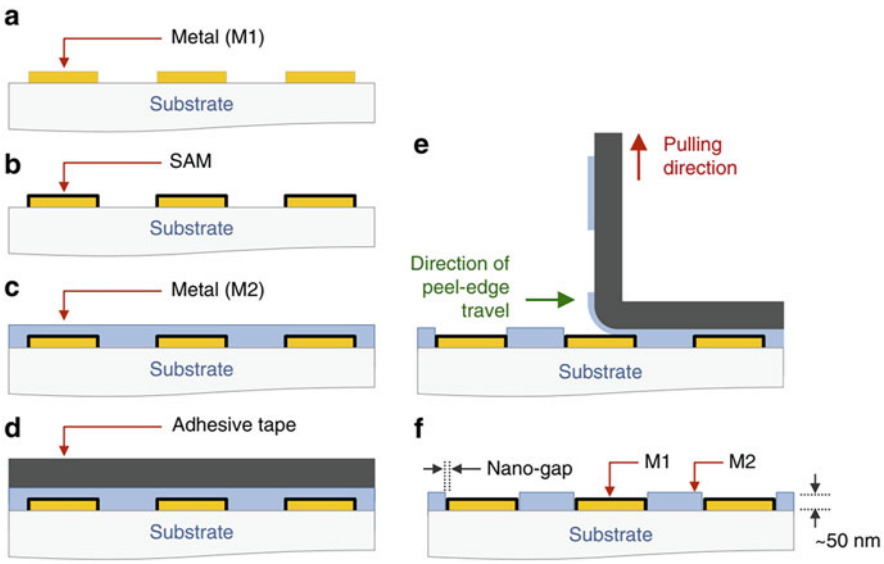


Fig. 9.4 Adhesion lithography to create nanogaps between metal electrodes (steps from **a** to **f** are described in the context)

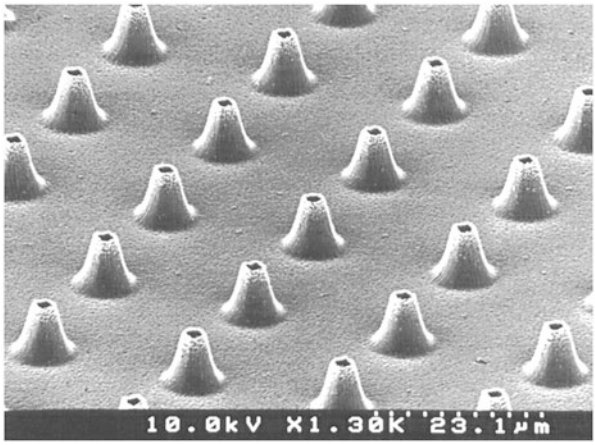


Fig. 9.5 SEM photo of metallic micronozzle array made by conformal coating and peel-off

substrate and metal deposition followed by peel-off using an adhesive tape, which have created the insulating gap as small as 1 nm [15]. Using adhesive tape to replace liftoff was also applied to make a variety of metallic patterns including sub-15 nm nanogaps in the so-called "sketch and peel" lithography [16]. Conformal metal coating on nonvertical silicon cones followed by peel-off was able to make metallic micronozzles as shown in Fig. 9.5 [17].

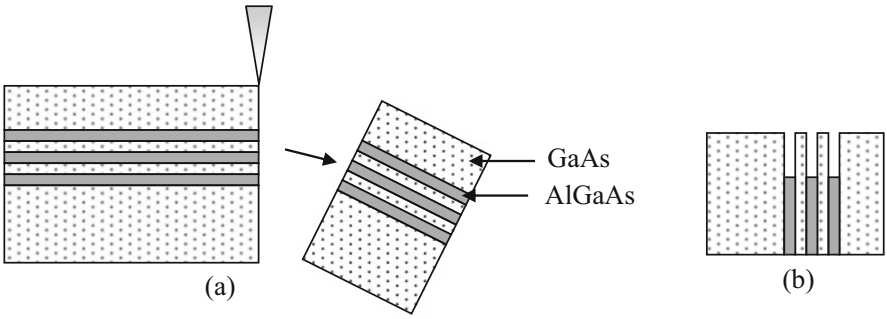
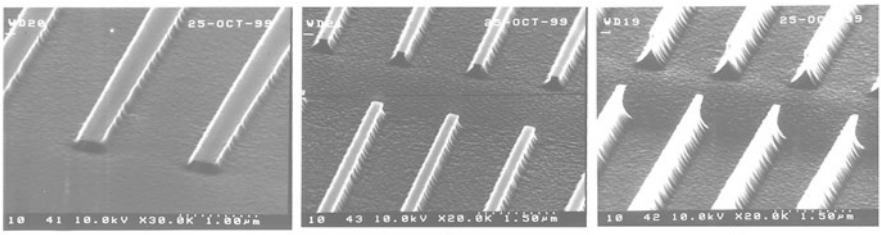


Fig. 9.6 Selective etching of MBE grown alternating layers: (**a**) MBE growth and sample cleaving; (**b**) selective etching

An extension of sidewall lithography is to planarly grow GaAs and AlGaAs alternative layers by molecular beam epitaxy (MBE) with precisely controlled thickness and then selectively etch away one of the materials, leaving the other as the final pattern structure. In order to get etchant solution into the buried layers, the substrate after growing the layers should be cleaved so that the alternating layer structures are exposed to wet chemical etchant, as schematically shown in Fig. 9.6a. Selective removal is possible because GaAs and AlGaAs have the property that they can be etched in different acidic solutions. For example, GaAs can be etched away in the mixture of citric acid and H_2O_2 (5:1 mixture) without affecting the $Al_{0.32}Ga_{0.68}As$ [18], or $Al_{0.7}Ga_{0.3}As$ can be etched away in a diluted solution of hydrofluoric acid without affecting the GaAs [19]. Acidic etching removes one of the crystalline materials and leaves the other alternating layers as the densely packed line pattern structure such as the one schematically shown in Fig. 9.6b. It is apparent that these line patterns will have the width equal to the thickness of MBE grown layers. A grating at 6 nm half-pitch was fabricated by cleaving AlGaAs superlattice grown on GaAs with molecular beam epitaxy and selectively etching away the AlGaAs layers in dilute hydrofluoric acid, which was then used as a nanoimprint mold [20].

9.3 Dimensional Subtraction

Taking away bit by bit from sideway of a wide line to make it narrow is the idea of indirect patterning by lateral subtraction. As introduced in Chap. 7, RIE always has certain degree of isotropic etching. While the etching is going deeper, it is also going wider laterally. The anisotropy can be controlled by tuning the plasma condition. Figure 9.7 shows a series of SEM photos of etched silicon stripes [21]. It demonstrates how the stripes are narrowed with the progression of etching time. Eventually, the structure dimension at the top can be made infinitely small. That is how very sharp silicon tips are made, as shown in Fig. 9.8 [22].



(a) (b) (c)

Fig. 9.7 Lateral subtraction of silicon line pattern in RIE: (**a**) etching 1 min; (**b**) etching 2 min; (**c**) etching 3 min

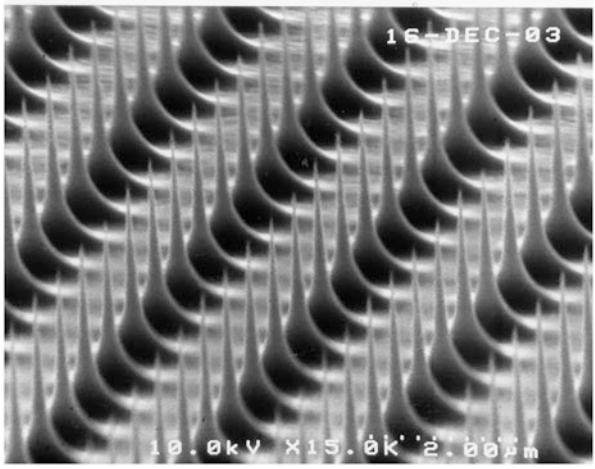
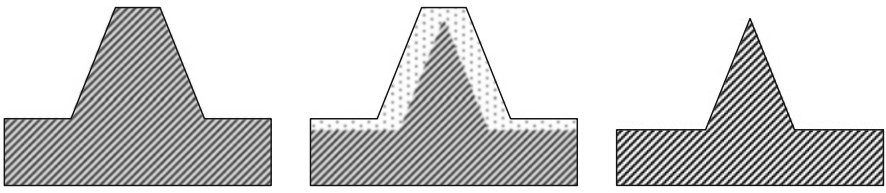


Fig. 9.8 Silicon tips made by RIE with nanometer scale apex



(a) (b) (c)

Fig. 9.9 Schematic of oxidation sharpening process: (**a**) initial tip geometry; (**b**) oxidation; (**c**) final tip geometry after removal of oxide layer

Another lateral subtraction technique, particularly suitable for scaling down silicon structures, is oxidation sharpening. Oxidation sharpening technique was first reported in 1991 [23]. It is known that silicon can be turned into silicon dioxide in thermal oxidation process. Once becoming oxide, it can be etched away in hydrogen fluoride (HF) acid. The process is schematically illustrated in Fig. 9.9.

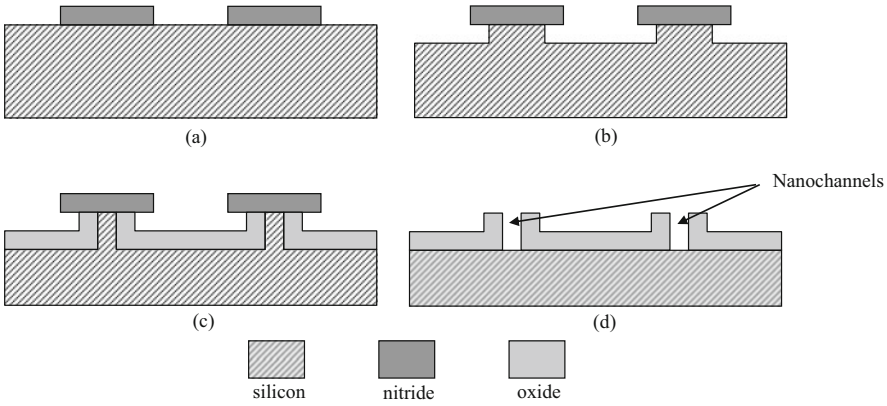


Fig. 9.10 Schematic of nanochannel fabrication by lateral oxidation and selective etching

An initial large silicon cone structure (Fig. 9.9a) is thermally oxidized. The thermal oxidation "consumes" away the silicon from surface down (Fig. 9.9b). The longer the thermal oxidation process, the more the silicon is converted into silicon dioxide. After the thermal oxidation, a HF wet etching removes the oxide "shell," leaving a sharpened silicon tip (Fig. 9.9c). Generally, a "fat" structure can be "slimmed" down by the oxidation and HF etching. It should be noted that the oxidation slows down with the oxidation layer growing thicker. Multiple oxidation-etching may be performed if significant amount of silicon has to be removed. Atomically sharp silicon tips, with a tip radius of <1 nm, have been made with two consecutive oxidation and HF etching processes [23]. Thermal oxidation requires temperature more than 950 °C. Alternatively, silicon substrate can be wet oxidized by immersion into HNO_3 solution. A saturated oxide layer about 1.8 nm thick can be formed in 5 min, which is then stripped away in HF solution. By repeated oxidation and HF etching, a fat silicon structure can be thinned down. In this way, 40 nm silicon lines were reduced to 10 nm after 15 cycles of oxidation and etching [24].

Instead of stripping oxide, the oxide layer can remain while the silicon is etched away, which results in holes or trenches whose dimension is determined by the amount of remaining silicon. The fabrication process is a combination of anisotropic etching in potassium hydroxide (KOH) solution, local oxidation of silicon (LOCOS), and plasma etching of silicon. Such a process has been utilized to fabricate sub-100 nm fluidic channels on a silicon substrate [25]. The process is schematically shown in Fig. 9.10. It starts from patterning Si_3N_4 mask of much wider dimension (2.5–3 μm) on a $<100>$ silicon substrate. This can be done by standard low-resolution contact photolithography. The lines are patterned parallel to $<110>$ directions (Fig. 9.10a). Once the Si_3N_4 lines are defined, wet chemical etching of silicon by KOH is performed at the bare silicon area (Fig. 9.10b). The next step is to oxidize the silicon by LOCOS at 1100 °C to grow SiO_2 to a thickness of more than 500 nm in both depth and lateral directions (Fig. 9.10c). The final step

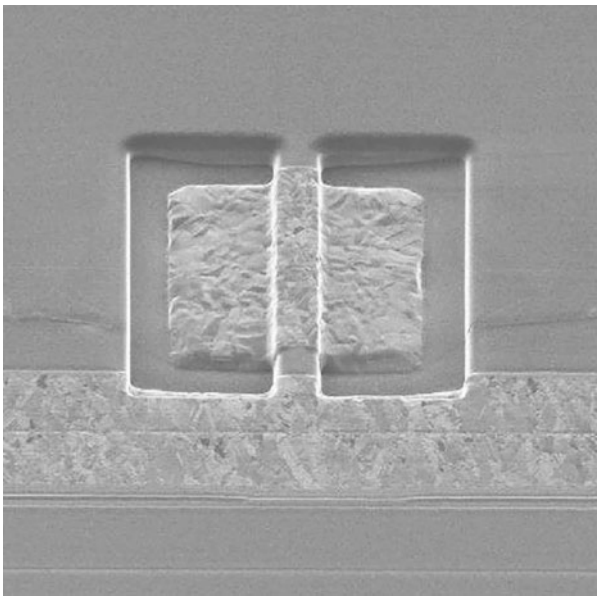


Fig. 9.11 Magnetic read/write head trimmed by FIB. (Reprinted from [27] with permission)

is to remove the Si_3N_4 mask and etch away the silicon sandwiched between the oxide walls (Fig. 9.10d). Thin channels as narrow as 40 nm can be fabricated without using any expensive lithography tools. This is a typical lateral subtraction process which takes advantage of the fact that silicon, nitride, and oxide can each be selectively etched without harming the others.

The lateral subtraction also has a name called "post-trimming." In fact, post-fabrication trimming is a widely used technique for feature size reduction [1]. For example, RIE has been used to trim electron beam exposed HSQ resist structure to achieve sub-10 nm pattern resolution. With CF_4 plasma trimming, an initial 21 nm HSQ line feature can be reduced to 5.3 nm [26]. Focused ion beam (FIB) is best suited to perform the trimming task, due to its precisely controlled scanning and sputter etching nature. FIB is flexible to sputter away materials from a very precise location. The best example of FIB trimming is the manufacturing of magnetic read/write heads, whereby the FIB trimming has become a standard volume manufacturing process. A magnetic read/write head initially patterned by optical lithography can be further trimmed down to sub-100 nm dimension by FIB sputter etching. Figure 9.11 is an example of FIB trimmed magnetic head [27]. An initial square-shaped pattern has been trimmed into a thin line.

There can be many clever ways to realize feature size reduction. Figure 9.12 illustrates how overexposure of positive-tone resist can produce much reduced size of negative-tone resist image. The standard mask pattern looks like what is shown in Fig. 9.12a. By overexposure, each dot expands laterally (white regions). After development, the remaining resist becomes the pattern of reversed tone with much reduced dimension (black dots inside white region), as shown in Fig. 9.12b. The size

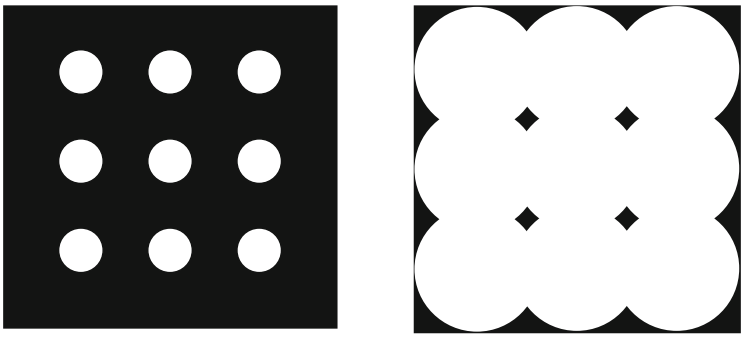


Fig. 9.12 Feature size reduction by overexposure of positive-tone resist: (**a**) initial positive-tone mask pattern; (**b**) final negative-tone pattern (four little dots in the middle) after overexposure of the pattern in (**a**)

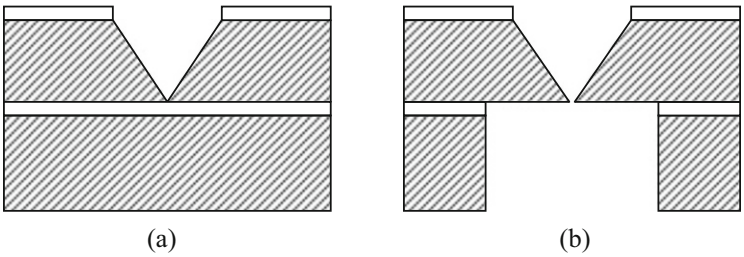


Fig. 9.13 Wet chemical etching of silicon to create nanoscale see-through hole: (**a**) reverse pyramid-shaped cavity is made in the top layer silicon of a SOI wafer; (**b**) removal of silicon and oxide from bottom side to reveal the hole

and period of reversed-tone dots array can be varied by varying the overexposure dose and the initial mask pattern design. It has been shown that for initial 125 nm positive-tone dot design, the exposure at 230 mJ/cm^2 produced 107 nm negative-tone dots, against the normal exposure dose of 110 mJ/cm^2. At 350 mJ/cm^2, the negative-tone dots were reduced to 53 nm [28].

Subtraction can also be performed vertically to create nanoscale see-through holes or slits. An example is to utilize anisotropic wet chemical etching of silicon. As described in Chap. 7, wet chemical etching of silicon in alkaline solution is dependent on crystalline orientation, i.e., etching fast in one direction but slow in another direction. A typical example is the etching of <100> silicon in KOH. The slow etching in <111> direction creates an etched profile with sloped sidewall. By controlling the mask opening size and etch depth, a reverse pyramid-shaped cavity can be created as shown in Fig. 9.13a. In this case, the reversed cone is made in a SOI substrate, and the etching is controlled to stop just when the tip of the reversed cone reaches the SiO_2 insulating layer. A second deep RIE of silicon from the bottom side

removes the silicon underneath the pyramid tip. A further etching of the SiO_2 layer reveals the cone tip which becomes a see-through hole, as shown in Fig. 9.13b. Square nanopore arrays with an average pore size of 60 nm as well as the smallest nanoslits of 13 nm have been made by this simple technique [29].

9.4 Dimensional Addition

While subtraction can produce narrow lines, the opposite is to produce narrow gaps by lateral addition. Conventional technique to create a narrow gap is direct e-beam lithography of resist and pattern transfer by RIE or direct FIB cutting of a continuous line. The lateral addition technique is to start with a wider gap and then narrow it by adding material bit by bit to close in the gap. The initial gap can be produced by conventional lithography and etching, as illustrated in Fig. 9.14a. Subsequent deposition of material closes the gap by hanging over the gap (Fig. 9.14b). By carefully controlling the initial gap distance and material deposition, 2–3 nm gap between two gold electrodes has been created [30]. The most flexible lateral addition process is to utilize electron beam or focused ion beam induced deposition, by which the amount of material deposited and the location of deposition can be precisely controlled. Figure 9.14c shows a sequence of process to produce a narrow gap by electron beam induced deposition of platinum. The initial 120 nm gap has been narrowed to 5 nm by gradual deposition of material from both sides [31].

Another additive method is to utilize the fact that oxidation can make materials grow "fatter," which can indirectly reduce the distance between two originally "slim" structures. Figure 9.15 illustrates such a scheme. The process starts with standard optical or e-beam lithography to pattern gold and chrome double-layer structure (Fig. 9.15a). The chrome is then allowed to oxidize to become CrO, which expands and overhangs on top of gold pads (Fig. 9.15b). Evaporation of gold blanketly covers the overhanging CrO pads and the rest of substrate (Fig. 9.15c). A liftoff to remove the CrO pads together with gold on top leaves the surface with gold patterns having gaps which are defined by the overhanging CrO structures (Fig. 9.15d). Using this nanomasking method, gold electrode gaps of 7.4 nm with a standard deviation of 2.2 nm have been produced [32].

Instead of oxidation, shrinking of lateral dimension can be realized by material deposition. As illustrated in Fig. 9.16, the initial holes or gaps made by optical lithography may be big (Fig. 9.16a). By depositing thin film, the diameter of the distance between adjacent patterns can be reduced, depending on the thickness of thin film (Fig. 9.16b). Plasma etching back removes the thin film from the top and bottom (Fig. 9.16c) but leaves the sidewall thin film intact. A subsequent pattern transfer by RIE results in much smaller holes or gaps in the substrate than original photoresist pattern (Fig. 9.16d). It is very much like the spacer concept shown in Fig. 9.1, but here the spacer acts as an etching mask instead of the pattern structure itself. A similar approach was used to make nanoscale sealed channels as illustrated in Fig. 9.17a. Conformal coating not only reduced the diameter of circular channel to

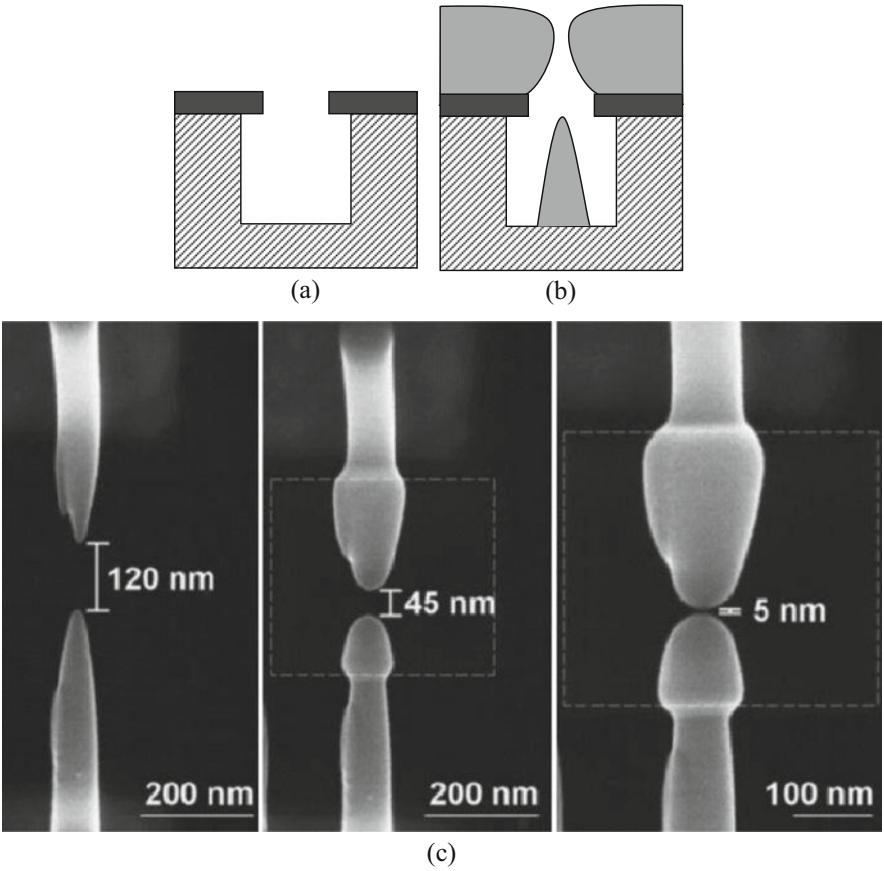


Fig. 9.14 (**a**) Initial gap produced by conventional lithography and etching. (**b**) Close the wide gap by material deposition. (**c**) An example of gap by electron beam induced Pt deposition. (Reprinted from [31] with permission)

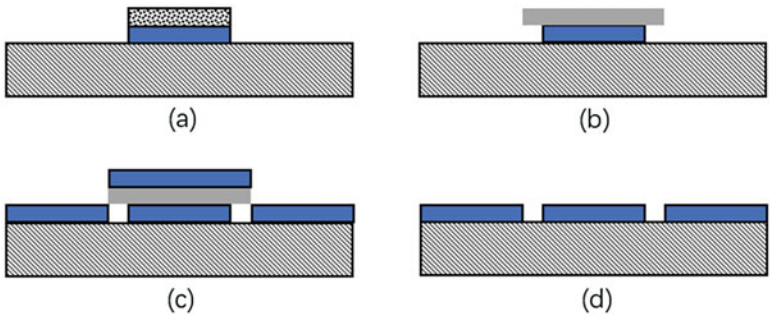


Fig. 9.15 Nanomasking process to produce gold electrode gaps: (**a**) initial Cr on top of Au pattern; (**b**) oxidization of Cr; (**c**) blanket coating of gold; (**d**) removal of CrO cap

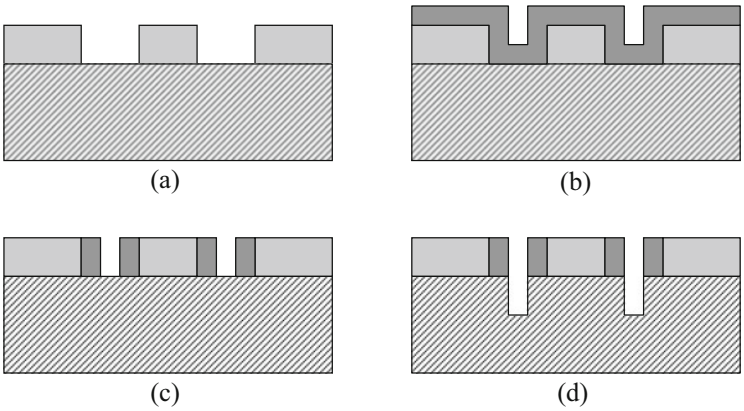


Fig. 9.16 Creation of nanoholes by lateral addition of material: (**a**) original photoresist pattern; (**b**) thin film deposition; (**c**) plasma etching back; (**d**) transfer to substrate

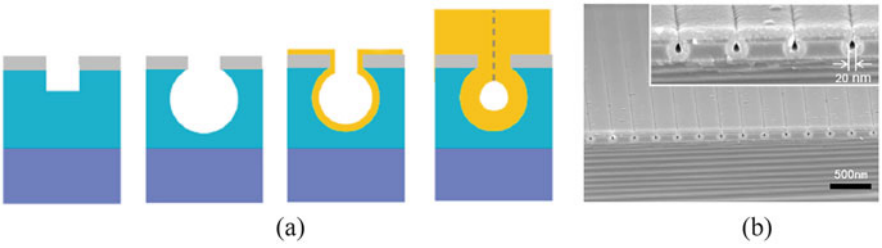


Fig. 9.17 (**a**) Circular channels with reduced diameter and sealed by ALD; (**b**) SEM image of fabricated nanoscale channels. (Reprinted from [33] with permission)

20 nm but also completely sealed it from top, as the SEM image shown in Fig. 9.17b [33].

It has been found that ion irradiation of a freestanding thin film can induce free movement of atoms in the thin film. If there is a hole in a freestanding thin film, ion beam irradiation of the hole and surrounding area can cause the hole to shrink, even close up completely. This so-called "ion beam sculpting" technique has been used to produce nanopores of less than 10 nm in thin membranes [34]. Conventionally, nanopores are made by RIE or direct FIB etching of a membrane. The direct technique is always limited by the resolution of lithography tool and process. The smallest hole etched by focused ion beam in a 30-nm-thick silicon nitride membrane is around 30 nm [35]. Further reduction of nanopore size beyond the resolution limit of FIB can be realized by the ion beam sculpting process. For example, an initial 100 nm hole in a 100-nm-thick silicon nitride membrane has been reduced to 15 nm after 18 scans of Ga ion beam in a 1–2 μm^2 area surrounding the hole. The complete process takes about 1 min. Figure 9.18a–d shows a series of SEM photos of a

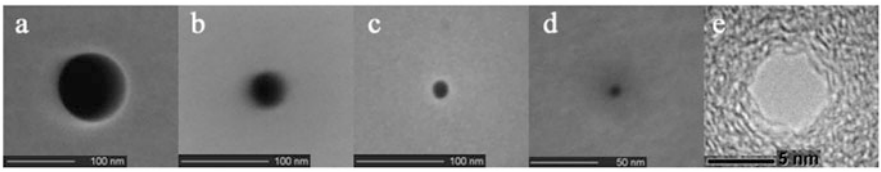


Fig. 9.18 Reduction of nanopore size by ion beam irradiation: (**a**) 100 nm, (**b**) 40 nm, (**c**) 20 nm, (**d**) 10 nm, (**e**) 5 nm (TEM image). (Reprinted from [35] with permission)

nanopore, representing gradual reduction of pore size from 100 nm to 5 nm through scanning of focused ion beam. The nanopore can be made even smaller if electrons instead of ions are used. By employing 200 keV electron beam in a transmission electron microscope (TEM), the 5 nm nanopore shown in Fig. 9.18e has been reduced to 1 nm [35]. The exact mechanism of beam induced nanopore shrink is still not clear. It is believed that high energy ions or electron irradiation of the thin membrane has increased the mobility and diffusivity of atoms surrounding a hole. The "melting" of material causes the close-up of nanopore. Nanopores can also be created by melting prepatterned polymeric materials because heating can induce reflow of polymers. Pore size as small as 6 nm was produced in SU-8 membrane by heating up the initially nanoimprinted holes in the polymer membrane [36].

9.5 Indirect Fabrication of Nanogaps

Nanogaps are a class of nanostructures which have drawn many interests from many fields of applications. In the last 6 years since 2018, there are more than 600 research papers published on the topic of nanogaps, more than 100 papers each year, according to the Web of Science database. Fabrication of nanogaps is also a research area where many ingenious indirect methods have been created, particularly the techniques on making metallic nanogaps [37]. The previous sections of this chapter already introduced some examples of indirect fabrication of nanogaps, such as the methods shown in Figs. 9.4, 9.10, 9.14, 9.15, and 9.16. Here are some more methods to realize extremely small gaps.

In addition to the sidewall lithography, dimensional subtraction or addition, electromigration effect has been utilized to create nanogaps whereby atoms migrate under the influence of electric field. When a conductive track is sufficiently narrow, electromigration can cause the broken track, which has been the major cause of interconnect failures in integrated circuits and other electronic devices [38]. However, this phenomenon can be used to make small gaps. 2–3 nm gaps between two electrodes have been made by this method [39]. FIB milling can also induce breaking of single grain boundary in polycrystalline material, resulting in nanogaps as narrow as 1–2 nm [40]. While the electromigration breaks a continuous electrode to create nanogaps, field emission current can reduce a large gap by accumulated

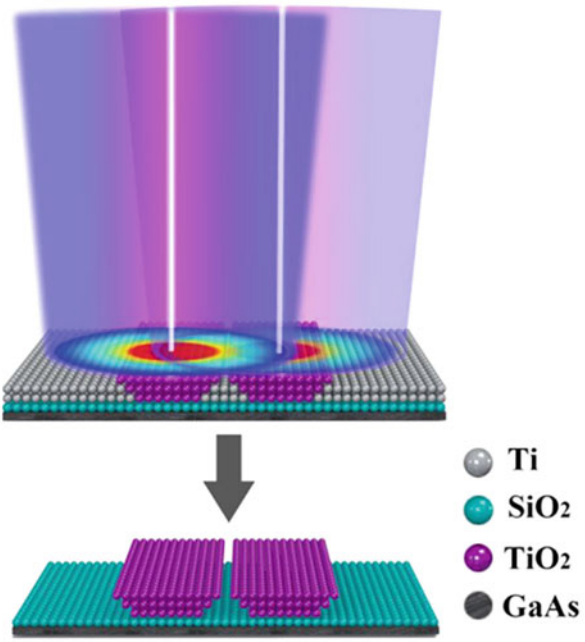


Fig. 9.19 Laser-induced oxidation of Ti and selective wet etching to create 5 nm nanogaps. (Reprinted from [43] with permission)

atomic migration. An initial gap separation of 50–71 nm was uniformly reduced to $<$10 nm by the accumulation of nickel atoms between two nickel tips at the field emission current of 300 nA [41]. Small gaps can also be produced by mechanically breaking a wire. Mechanical cracking is carried out with stress concentration in prepatterned notches and necks defined by lithographic patterning [1]. For example, thin gold lines were made on a silicon wafer, and the silicon wafer was mechanically bended to generate a tensile stress in the gold lines. A breakage occurred at the narrowest part the line and made gaps as narrow as 0.8 nm [42].

While a nanogap can be produced by deposition from both sides of a large gap as seen in Fig. 9.14, it can also be created by modifying the material property from both sides of a large gap and subsequently selective etching to remove the middle part. Figure 9.19 illustrates such a scheme based on laser-induced thermal oxidation effect [43]. Instead of exposing photoresist by a laser beam, titanium (Ti) is used as the imaging layer which can be oxidized by laser-induced thermal effect. The threshold energy for laser oxidation of Ti is about 127.3 J/cm^2 [44]. When scanning with two laser beams close enough, Ti on the scanning path is oxidized but leaves a gap where Ti is not oxidized because there is an energy "valley" which is below the oxidation threshold. A nanogap is then obtained by wet etching to remove Ti in the middle because of the high contrast selective etching rate between Ti and TiO_2 with the etchant of HF: (25%)/H_2O_2 (30%)/H_2O = 1:1.5:100 [43].

9.6 Super-resolution Patterning

It has been introduced in Chap. 2 that the resolution of optical lithography is limited by diffraction. Super-resolution is the imaging technique that can break or overcome the diffraction limit, which is also called subdiffraction techniques. If closely examining the light intensity distribution shown in Fig. 2.4 of Chap. 2, one can see that the intensity shape is always narrow at the top and broadened downward. If the exposure of photoresist is only confined to the top section of the intensity, then the obtained resolution would be much higher. One of the developments to achieve super-resolution has been the two-color photolithography [45]. The basic idea is to use two wavelengths (two colors) of photon beams. One wavelength beam performs normal photoresist exposure (activation), and the other wavelength beam stops the photoresist exposure (inhibition). By surrounding the activation beam with the inhibition beam, the exposure is confined to only very high intensity part which is of much narrow distribution, hence very high resolution.

There are two ways to realize such scheme. One is to use an absorbance modulation layer (AML) as illustrated in Fig. 9.20. The AML is a photochromic layer which is initially transparent but becomes opaque when illuminated by λ_2 wavelength beam. The photoresist underneath the AML only responds to λ_1 wavelength beam. The λ_2 beam exposure makes the transparent opening of AML only at the nodal point which allows only the peak point of λ_1 beam intensity distribution to expose the photoresist underneath. Super-resolution optical lithography is then achieved. With $\lambda_1 = 325$ nm and $\lambda_2 = 633$ nm, minimum 30 nm line features were produced in PMMA resist, which is only one-tenth of the illuminating wavelength [46]. The other way to realize two-color photolithography is to formulate the photoresist whose polymerization is inhibited at one wavelength but activated at another wavelength. With a femtosecond laser ($\lambda_1 = 800$ nm) as the activation beam

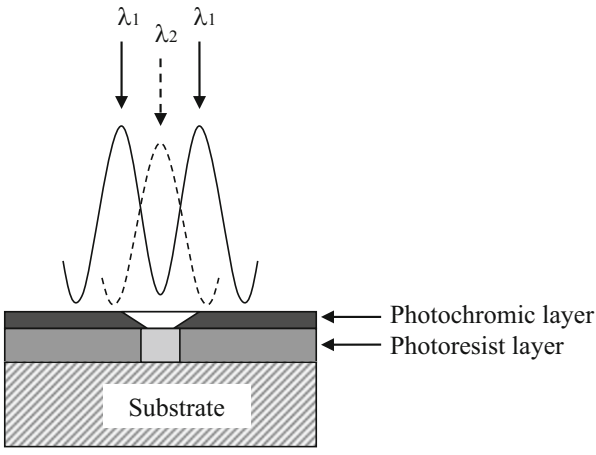


Fig. 9.20 Super-resolution optical lithography enabled by absorbance modulation layer

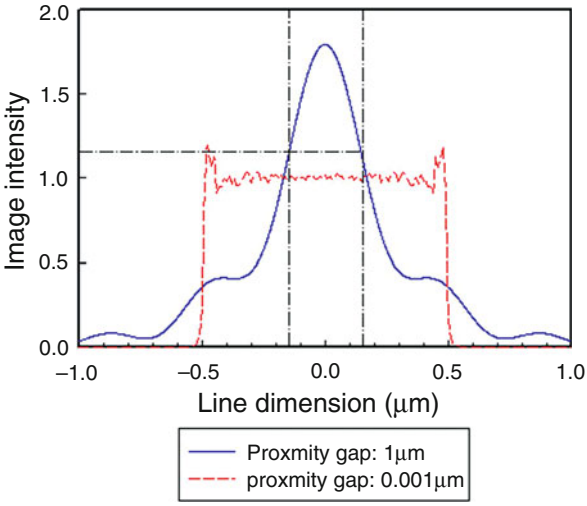


Fig. 9.21 Light intensity distribution on photoresist surface at different mask-substrate distances

and a doughnut-shaped continuous laser ($\lambda_2 = 375$ nm) as the inhibition beam, the smallest line feature of 9 nm and two-line pitch of 52 nm were achieved [47]. Recently, the method was employed to make 3D optical disk with petabit (10^{15} bits) storage capacity [48].

In proximity optical lithography, i.e., 1:1 optical imaging by a mask aligner, there is an interesting phenomenon that the light intensity distribution changes its shape with changing gap distance between mask and substrate. Figure 9.21 shows the simulated light intensity distributions at two different gap distances between mask and substrate (photoresist surface). If the gap distance is infinitely small (assumed 0.001 μm in simulation), a square-shaped light distribution has almost the same width as the mask feature opening (1 μm). When the mask moves away from the photoresist surface, the light distribution will change due to diffraction. However, there is a gap distance (1 μm in this case) at which the top section of the distribution is much narrower than the opening width of mask feature. Further increasing the gap distance, the distribution becomes flat and the high resolution peak disappears. If the mask-substrate gap is controlled at this optimum distance and the exposure dose is controlled to enable only the top section of light intensity to participate the photoresist exposure, then one can expect the photoresist image to be much smaller than the mask feature width. In the simulation example shown in Fig. 9.21, the line feature, obtained by separating the mask from photoresist surface with 1 μm gap and utilizing only the top part of light intensity, will be only 300 nm instead of 1 μm of the mask design.

Such phenomenon has been utilized in X-ray lithography, since X-ray lithography is also a proximity imaging [49]. Figure 9.22 compares the SEM image of X-ray mask features which is measured to be 152 nm line opening (Fig. 9.22a) and exposed line image in SAL605 negative-tone resist which is measured to be 61 nm

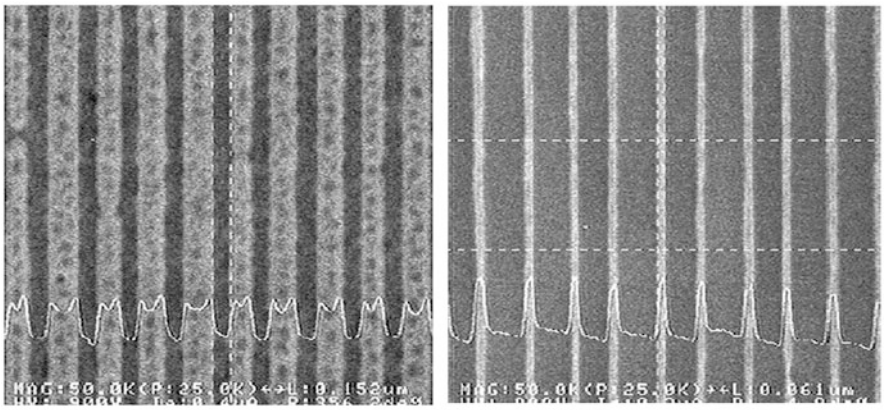


Fig. 9.22 Demagnification of X-ray proximity printing: (**a**) mask feature at 152 nm opening width; (**b**) exposure resist feature at 61 nm line width. (Reprinted from [49] with permission)

(Fig. 9.22b). The demagnification is around 2.5. As aforementioned, to produce high resolution patterns by a low-resolution mask in proximity X-ray lithography, not only should the mask-substrate gap distance be carefully controlled and optimized but also the careful control of exposure dose and development time. The 61 nm lines in SAL605 negative-tone resist were obtained by 1 min 45 s development time. The same resist feature became 43 ~ 46 nm if developed for 2 min [49]. The longer development time has tailored the exposure level toward higher threshold intensity at which the distribution is narrower. A simulated example is shown in Fig. 9.23 where the intensity profile is generated by optical interference lithography. By controlling the threshold intensity at 1.5, the exposed lines can be much narrower.

Controlling exposure dose and development time to acquire different exposed feature dimension in resists is also a common practice in electron beam lithography. It is a well-known fact that the highest resolution e-beam lithography is only achievable in resists having the lowest sensitivity such as PMMA, HSQ, and some inorganic resists. The reason is that an electron beam has a Gaussian distribution both in beam current and in electron energy and that only the high energy electrons and peak intensity of beam current, which is very narrow in distribution, can expose a low sensitivity resist. The resultant feature will be narrow because the broad portion of the beam, which mainly consists of low energy scattering electrons and low number of electrons, will have negligible effect on the polymer molecular chains in the resist. This also true for laser direct write lithography because a laser beam also has a Gaussian-shaped energy distribution. By laser-induced thermal oxidation of Ti film, which is of much lower sensitivity to laser energy than exposure of photoresists, 40 nm feature size was produced by a 532 nm wavelength laser [44].

The same practice can be applied to focused ion beam (FIB) as well. Although the beam current distribution in FIB is generally Gaussian like, it has a larger tail than an electron beam because of the high energy spread in a gallium liquid metal ion source.

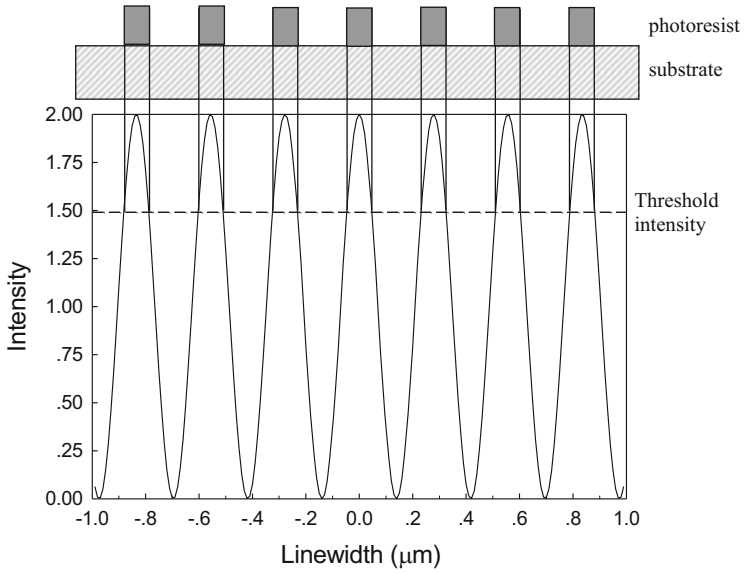


Fig. 9.23 Dependance of exposed linewidth on threshold intensity level

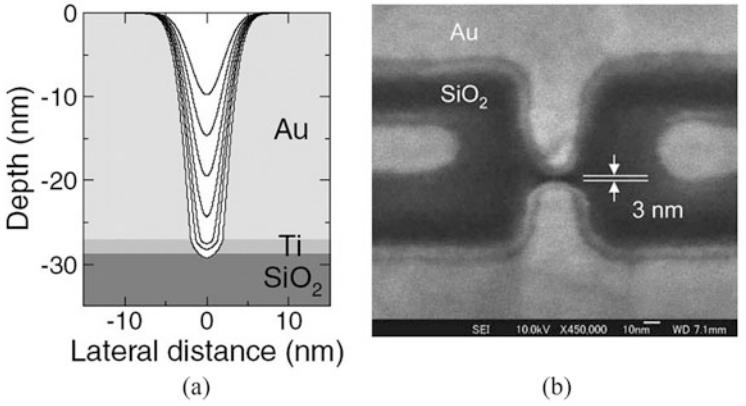


Fig. 9.24 (**a**) Simulated contours of FIB sputtering of Au-Ti film; (**b**) SEM image of 3 nm nanogap created by FIB sputtering with beam diameter of ~12 nm. (Reprinted from [50] with permission)

This is one of the reasons that the resolution of patterning by FIB is generally not as high as that by an electron beam. However, by utilizing the beam profile, it is possible to fabricate structures much smaller than the beam size. Figure 9.24 shows how a narrow gap can be made by FIB sputter etching [50]. The computer simulated contours of FIB sputtering in Fig. 9.24a represent the successive sputtering profiles in the Au film as the FIB sputtering dose increases (Au film of 27 nm thickness on top of SiO_2 with a 1.6 nm Ti seed layer). There is a sputtering

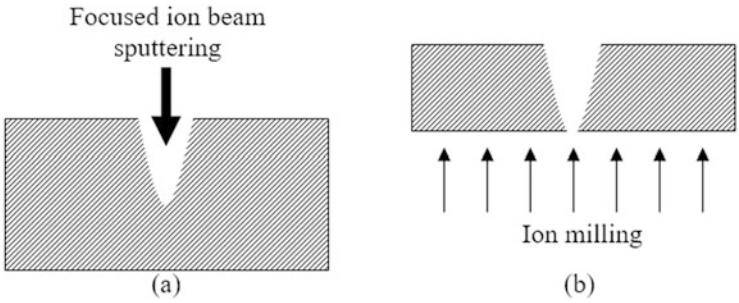


Fig. 9.25 Nanopore produced by FIB sputtering and ion milling: (**a**) FIB sputtering from top side of a membrane; (**b**) ion milling from bottom side of membrane

dose at which the Au-Ti film is just broken through by the FIB etching and a narrow gap is created in the film. Figure 9.24b is the SEM image of an actual 3 nm nanogap created by FIB sputtering with a beam diameter of ~12 nm. To produce this nanogap, it is very critical to be able to control the FIB sputtering to stop just at the moment when the beam tip breaks through the Au-Ti film. To do this, an in situ measurement of current flowing through the electrode is performed, which is connected to the FIB scan controller. The beam sputtering scan is switched off immediately when the current flow through the thin film electrode drops to zero, which indicates that the electrode has been cut through by FIB sputtering.

The Gaussian distribution of a FIB determines that a hole sputtered by FIB will also be in the shape of a Gaussian profile, as shown in Fig. 9.25a. Then taking away materials layer by layer from the other side would eventually reveal the tip of this Gaussian-shaped hole and create a very small pore which diameter is determined by the Gaussian distribution width as illustrated in Fig. 9.25b. The key step in making the nanopore is the precise control of ion milling depth in the backside. The ion milling has to stop just as the membrane has been broken through. To achieve this precise control, an ion counter can be mounted on the other side of membrane. When ions start to flow through and are detected by the ion counter, it signals the breakthrough of the membrane and the time to stop the ion milling. The size of nanopore can be further reduced by low dose exposure of ions after a hole is created, which promotes the mobility of atoms surrounding the hole and shrinks the hole, i.e., the ion beam sculpting mentioned previously. Nanopore with diameter as small as 1.8 nm have been created in Si_3N_4 membrane by the technique [34].

References

1. Chen, Y., et al.: Sub-10 nm fabrication: methods and applications. Int. J. Extrem. Manuf. **3**, 032002 (2021)
2. Sakamoto, J., et al.: High aspect ratio nano mold fabrication by advanced edge lithography without CVD. Microelectron. Eng. **88**, 1992–1996 (2011)

3. Flanders, D.C., Efremow, N.N.: Generation of <50 nm period gratings using edge defined techniques. J. Vac. Sci. Technol. B. **1**, 1105–1108 (1983)
4. Vettiger, P., et al.: Nanometer sidewall lithography by resist silylation. J. Vac. Sci. Technol. **B7**(6), 1756 (1989)
5. Cai, M., et al.: Efficient annular aperture array (AAA) filter for thermophotovoltaic by sidewall lithography. J. Micromech. Microeng. **32**, 085008 (2022)
6. Lo, Y.-C., Cheng, X.: Polymeric sidewall transfer lithography. J. Phys. Commun. **6**, 095007 (2022)
7. Hsu, D.S.Y.: 50 nm. Linewidth ~platinum sidewall Aithography by effusive-source metal precuisor chemical. Deposition and ion-assisted etching. Appl. Phys. Lett. **59**(17), 2192 (1991)
8. Yu, Z., et al.: Fabrication of large area 100 nm pitch grating by spatial frequency doubling and nanoimprint lithography for subwavelength optical applications. J. Vac. Sci. Technol. **B19**(6), 2816 (2001)
9. Choi, Y.-K., et al.: Fabrication of Sub-10-nm silicon nanowire arrays by size reduction lithography. J. Phys. Chem. B. **107**, 3340–3343 (2003)
10. Zaborowski, M., et al.: Nano-line width control and standards using lateral pattern definition technique. Microelectron. Eng. **83**, 1555–1558 (2006)
11. Raley, A., et al.: Self-aligned quadruple patterning integration using spacer on spacer pitch splitting at the resist level for sub-32nm pitch applications. Proc. SPIE. **9782**, 97820F (2016)
12. Owa, S., et al.: Immersion lithography extension to sub-10 nm nodes with multiple patterning. Proc. SPIE. **9052**, 90520O-1 (2014)
13. Yaegashi, H., et al.: Enabled scaling capability with self-aligned multiple patterning process. J. Photopolym. Sci. Technol. **27**(4), 491–496 (2014)
14. Beesley, D.J., et al.: Sub-15-nm patterning of asymmetric metal electrodes and devices by adhesion lithography. Nat. Commun. **5**, 3933 (2014)
15. Chen, X., et al.: Atomic layer lithography of wafer-scale nanogap arrays for extreme confinement of electromagnetic waves. Nat. Commun. **4**, 2361 (2013)
16. Chen, Y., et al.: "Sketch and Peel" lithography for high-resolution multiscale patterning. Nano Lett. **16**, 3253–3259 (2016)
17. Cui, Z.: Fabrication of Nozzle array by surface moulding technique. In: International Microprocesses and Nanotechnology Conference, Digest of Papers, p. 94, Tokyo (2002)
18. Luber, S.M., et al.: Nanometre spaced electrodes on a cleaved AlGaAs surface. Nanotechnology. **16**, 1182–1185 (2005)
19. Austin, M.D., et al.: Fabrication of 5 nm linewidth and 14 nm pitch features by nanoimprint lithography. Appl. Phys. Lett. **84**(26), 5299 (2004)
20. Austin, M.D., et al.: 6 nm half-pitch lines and 0.04μm static random access memory patterns by nanoimprint lithography. Nanotechnology. **16**(8), 1058–1061 (2005)
21. Cui, Z., et al.: Fabrication of magnetic rings for high density memory devices. Microelectron. Eng. **61-62**, 577–583 (2002)
22. Tao, J., et al.: A systematic study of dry etch process for profile control of silicon tips. Microelectron. Eng. **78-79**, 147–151 (2005)
23. Ravi, T.S., et al.: Oxidation sharpening of silicon tips. J. Vac. Sci. Technol. B. **9**(6), 2733–2737 (1991)
24. Jo, M.-S., et al.: Realization of large-scale sub-10 nm nanogratings using a repetitive wet-chemical oxidation and etching technique. Micro Nano Syst Lett. **5**, 19 (2017)
25. Cho, Y.H., et al.: Fabrication of silicon dioxide submicron channels without nanolithography for single biomolecule detection. Nanotechnology. **18**, 465303 (2007)
26. Baek, I.-B., Yang, J.-H., Cho, W.-J.: Electron beam lithography patterning of sub-10 nm line using hydrogen silsesquioxane for nanoscale device applications. J. Vac. Sci. Technol. **B23**(6), 3120 (2005)
27. Reyntjens, S., Puers, R.: A review of focused ion beam applications in microsystem technology. J. Micromech. Microeng. **11**, 287 (2001)

28. Luo, Y., Misra, V.: Large-area long-range ordered anisotropic magnetic nanostructure fabrication by photolithography. Nanotechnology. **17**, 4909–4911 (2006)
29. Deng, T., et al.: Controllable fabrication of pyramidal silicon nanopore arrays and Nanoslits for nanostencil lithography. J. Phys. Chem. C. **118**, 18110–18115 (2014)
30. Steinmann, P., Weaver, J.M.R.: Fabrication of sub-5 nm gaps between metallic electrodes using conventional lithographic techniques. J. Vac. Sci. Technol. **B22**(6), 3178 (2004)
31. Gazzadi, G.C., et al.: Fabrication of 5nm gap pillar electrodes by electron-beam Pt deposition. J. Vac. Sci. Technol. **B23**(2), L1 (2005)
32. Bauman, S.J., et al.: Fabrication and analysis of metallic nanoslit structures: advancements in the nanomasking method. J. Micro/Nanolith. MEMS MOEMS. **17**(1), 013501 (2018)
33. Nam, S.-W., et al.: Sub-10-nm nanochannels by self-sealing and self-limiting atomic layer deposition. Nano Lett. **10**, 3324–3329 (2010)
34. Li, J., et al.: Ion-beam sculpting at nanometre length scales. Nature. **412**, 166 (2001)
35. Lo, C.J., Aref, T., Bezryadin, A.: Fabrication of symmetric sub-5 nm nanopores using focused ion and electron beams. Nanotechnology. **17**, 3264–3267 (2006)
36. Choi, J., et al.: Scalable fabrication of sub-10 nm polymer nanopores for DNA analysis. Microsyst. Nanoeng. **5**, 12 (2019)
37. Yang, Y., et al.: Sub-5 nm metal nanogaps: physical properties, fabrication methods, and device applications. Small. **15**, 1804177 (2019)
38. Kim, C.-U. (ed.): Electromigration in Thin Films and Electronic Devices: Materials and Reliability. Woodhead Publishing (2011)
39. Saha, S., Qian, G., Lewis, K.M.: Fabrication of nanogaps by a progressive electromigration technique using wires of various thicknesses. J. Vac. Sci. Technol. B. **29**(6), 061802–061801 (2011)
40. Cui, A., et al.: Single grain boundary break junction for suspended nanogap electrodes with gapwidth down to 1–2 nm by focused ion beam milling. Adv. Mater. **27**, 3002–3006 (2015)
41. Ito, M., et al.: High-throughput nanogap formation by field-emission-induced electromigration. J. Vac. Sci. Technol. B. **33**(5), 051801–051801 (2015)
42. Tsutsui, M., et al.: Atomically controlled fabrications of subnanometer scale electrode gaps. J. Appl. Phys. **108**, 064312 (2010)
43. Qin, L., et al.: 5 nm nanogap electrodes and arrays by a super-resolution laser lithography. Nano Lett. **20**, 4916–4923 (2020)
44. Wang, Y., et al.: Controllable fabrication of super-resolution nanocrater arrays by laser direct writing. J. Nanosci. Nanotechnol. **10**, 7134–7137 (2010)
45. Liaros, N., et al.: Ten years of two-color photolithography. Opt. Mater. Express. **9**(7), 3006–3020 (2019)
46. Andrew, T.L., et al.: Confining Light to DeepSubwavelength Dimensionsto Enable Optical Nanopatterning, vol. 324, pp. 917–920. Science (2009)
47. Gan, Z., et al.: Three-dimensional deep sub-diffraction optical beam lithography with 9 nm feature size. Nat. Commun. **4**, 2061 (2013)
48. Zhao, M., et al.: A 3D nanoscale optical disk memory with petabit capacity. Nature. **626**, 772–778 (2024)
49. Vladimirsky, Y., et al.: Demagnification in proximity x-ray lithography and extensibility to 25 nm by optimizing Fresnel diffraction. J. Phys. D. Appl. Phys. **32**, L114–L118 (1999)
50. Nagase, T., et al.: Maskless fabrication of nanoelectrode structures with nanogaps by using Ga focused ion beams. Microelectron. Eng. **78-79**, 253–259 (2005)

Chapter 10
Nanofabrication by Self-Assembly



10.1 Introduction

The nanofabrication technologies described in the previous chapters so far can be characterized as conventional technologies, in the sense that they always involve one way or another lithographic patterning and pattern transfer. They are still the same as, or not too much different from, the technologies for manufacturing integrated circuits, which originated more than 60 years ago when the first patent on planar integrated circuit was granted, though today's technologies can make pattern structures a thousand times smaller. It is a "top-down" approach that complex structures are built up by patterning layers upon layers from the surface of a planar substrate. The capabilities of top-down nanofabrication technologies have been amply demonstrated in previous chapters. The smallest features which can be made by top-down approach are, however, always limited by the available fabrication tools, being either lithography or pattern transfer. On the other hand, nature has demonstrated its ability to produce extremely complex living organisms by self-organization and self-construction for billions of years, all of which are taking place at molecular level which is an order of magnitude smaller than a nanometer. One surely could ask the question why a desired nanostructure cannot be made by molecular self-assembly, which is a "bottom-up" approach where molecules autonomously "grow" into nanostructures.

Such an attempt had started more than 30 years ago, pioneered by, not surprisingly, chemists and biologists [1]. To them, manipulation of molecules is a familiar subject. In some areas, nanostructures constructed by molecular self-assembly have been quite successful. For example, self-assembled monolayers (SAMs) have been widely used in nanoscale patterning. Creation and application of carbon nanotubes (CNT) is another example how nanostructures and objects can be grown in controllable manner. DNA molecular strands have been used as basic scaffolds to construct nanostructures. Controlled aggregation of colloidal particles is being widely utilized in formation of functional patterns and structures. There are other examples which

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_10

demonstrate certain 2D or 3D pattern structures that can be formed by guided self-assembly or self-organization from the interaction of molecular force or other interparticle forces.

The nanofabrication by self-assembly is inherently advantageous over conventional top-down technologies in two aspects: one is the ability to construct structures in true nanoscale, i.e., at molecule level or at least below 10 nm at which current patterning technologies are struggling or unable to achieve; another is to make nanostructures cheaply. Nanostructuring by molecular self-assembly should be, in principle, low cost, because there are no complicated and expensive fabrication equipment involved. For self-assembly to replace current lithography processes, the grand challenge is how to control the synthesis and assembly of molecules into desired nanostructures. As Professor George M. Whitesides of Harvard University put it in 1995, "Although there are countless examples of self-assembly all around us—from molecular crystals to mammals—the basic rules that govern these assemblies are not understood in useful detail, and self-assembling processes cannot, in general, be designed and carried out 'to order.' Many of the ideas that are crucial to the development of this area, 'molecular shape,' the interplay between enthalpy and entropy, the nature of non-covalent forces that connect the particles in self-assembled molecular aggregates, are simply not yet under the control of investigators" [2]. Gladly, such uncertainty for self-assembly to be one of the nanofabrication technologies has largely been eliminated after 30 years of continuous research efforts. Self-assembly processes can be designed and carried out to order.

One has to be aware that the current level of nanofabrication by self-assembly is still very primitive. Most of the nanostructures made from self-assembly are array like, which is originated from the formation of crystal lattices by molecular attractions. In majority of cases, the self-assembly has to be guided or to take place on a template, which requires assistance from conventional top-down nanofabrication processes. The combination of top-down and bottom-up approaches, or the so-called top-down meets bottom-up, has been, and will be in the future, the main feature in the development of self-assembly nanofabrication.

10.2 Self-Assembly Processes

Self-assembly has a very broad meaning. Any spontaneous formation of an organization from individual parts without human intervention can be called self-assembly. Self-assembly can take place either at molecular level or meso−/macroscopic level, from the formation of crystals to snowflakes and to universe. There are a number of distinctive features in self-assembly process:

1. Self-assembly is spontaneous, automatic, and autonomous. Autonomous means that once the conditions for self-assembly to occur are satisfied, the assembly process starts automatically and spontaneously without human intervention.

2. Self-assembly is the result of interaction of interfacial forces between individual entities. At molecular level, the interfacial forces involve non-covalent bonding force, typically van der Waals force, or weak covalent bonding force such as hydrogen bonds. At mesoscopic level, self-assembly may involve electrostatic and surface tension forces. In macroscopic scale, gravitational force, capillary force, or external electric or magnetic force may come into play [3].
3. The individual entities taking part in a self-assembly process have to be mobile in order for the self-assembly process to occur. Therefore, self-assembly process normally takes place in liquid phase or on a smooth surface so that the basic building blocks can move around and interact with each other.
4. The formation of an ordered structure is usually the result of equilibrium state, at which individual parts keep equal distance to each other, not a random aggregation. The self-assembled structures are long range ordered beyond a few molecules once the interaction reaches equilibration.
5. The actual formation of an organization or ordered structure is influenced by environment, either the input of external energy or confined geometrical boundaries.

A self-assembly process can be static or dynamic depending on if there are external influences. In the absence of external influences, building blocks self-assemble statically, driven by energy minimization to form static equilibrium structures. In the presence of outside influences, a dynamically self-assembling system may prevail that can adjust to its surrounding environment by residing on an energetic minimum which is caused by the influx of energy in the system. It reduces entropy by absorbing energy from the environment. This gradient in entropy between the organism and the environment can be maintained only as long as energy is driven from the environment into the organism [4]. A familiar example about how environment influences the self-assembly process is the formation of snowflakes. Water molecules crystallize from liquid phase to solid phase when temperature is below 0°C. The ice crystals are of many different shapes, depending on temperature, pressure, air current, and humidity, as the examples shown in Fig. 10.1. Generally, colder temperatures produce snowflakes with sharper tips on the sides of the crystals and may lead to more branching of the snowflake arms (dendrites). Snowflakes that grow under warmer conditions grow more slowly, resulting in smoother, less intricate shapes.

10.3 Self-Assembly of Atoms and Molecules

Self-assembly can take place at either atomic level or molecular level. There are typically two routes to realize atomic or molecular self-assembly: vapor phase or solution phase [6]. Most thin-film deposition processes encountered in nanofabrication can be characterized as vapor phase self-assembly, such as physical or chemical vapor depositions (PVD or CVD), which allows atomic species to



Fig. 10.1 Various shapes of snowflakes from crystallization of water molecules. (Reprinted from [5] with permission)

deposit on a surface where they attain mobility at certain temperature to assemble into crystal-like structures. They can also assemble through vapor phase in free space to form 3D nanostructures. The typical examples are how the carbon nanotubes and C_{60} are formed. The vapor phase self-assembly tends to form covalent bonds between atoms, which requires input of high energy from environment.

The self-assembly processes that are of interests in nanostructure formation are via solution phase. Highly ordered nanostructures can be fabricated by the spontaneous arrangement of atoms, molecules, colloids, or molecular units in solution or liquid phase into ordered structures via local interactions (e.g., electrostatic, van der Waals, hydrophobic, etc.) between the molecular units. Solution-based self-assembly methods offer numerous benefits for fabricating nanostructures such as potential for tunability of nanostructure properties by modifying concentrations of constituent molecules and ease of controlling shape and size of nanostructures. These self-assembled structures are easy to prepare at room temperature and form quickly and spontaneously from solutions of molecules. They are molecularly ordered and robust under many conditions of use. They are thermodynamically stable and tend to reject defects. They permit control of layer thickness to within approximately 1 nm [7].

10.3.1 Self-Assembled Monolayer

Solution-phase self-assembly of molecules was discovered more than a hundred years ago with the creation of Langmuir-Blodgett (LB) films [8]. While the LB films are still widely used in biological applications, the most successful molecular self-assembly in nanofabrication is the formation of self-assembled monolayer (SAM)

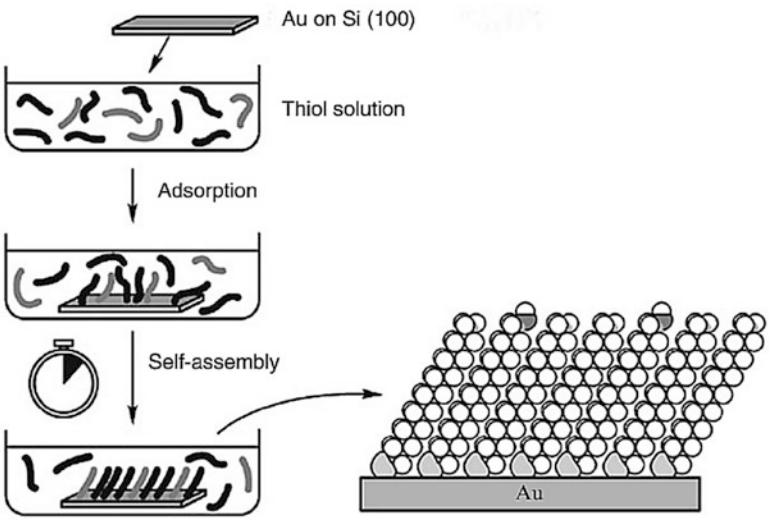


Fig. 10.2 Schematic of self-assembling monolayer on gold

that exhibited higher mechanical resistance and chemical stability than LB films [9]. SAMs are ordered molecular assemblies formed by the adsorption of an active surfactant on a solid surface. The order in these two-dimensional systems is produced by a spontaneous chemical synthesis at the interface, as the system approaches equilibrium. SAMs are usually prepared by immersing a substrate in the solution containing a ligand that is reactive toward the surface or alternatively by exposing the substrate to the vapor of the reactive species. The most investigated SAMs are alkanethiolates $CH_3(CH_2)_nS^-$ on Au(111) surface [10]. Alkanethiols absorb spontaneously on a gold surface from solution and form adsorbed alkanethiolates. Though the organosulfur compounds also absorb very strongly to silver, copper, platinum, mercury, iron, nanosized γ-Fe_2O_3 particles, colloidal gold particles, GaAs, and InP surfaces [9], the SAM on gold is of the best quality and mostly investigated and used. Figure 10.2 illustrates the SAM formation process.

The first step in creating the SAM is the absorption process where the sulfur atom chemisorbs to the Au(111). This is possible since sulfur compounds have a strong affinity to transition metal surfaces, such as Au, due to their ability to form multiple bonds with the surface metal clusters. Sulfur atoms bonded to the gold surface bring the alkyl chains into close contact; these contacts freeze out configurational entropy and lead to an ordered structure. The kinetics that drive this ordering are related to chain disorder, the dipole-dipole and van der Waals forces of the chain-chain interaction, and the surface mobility of the chains. The kinetics is faster for longer alkane chains, probably due to the increased van der Waals interactions. Highly ordered SAMs of hexadecanethiolate on gold can be formed in a few minutes by immersing a gold substrate in a solution of hexadecanethiol in ethanol (ca. 2 mM). In microcontact printing using a soft stamp, as described in Chap. 6, the transfer of thiol molecules onto a gold surface to become SAM takes only a few seconds [10].

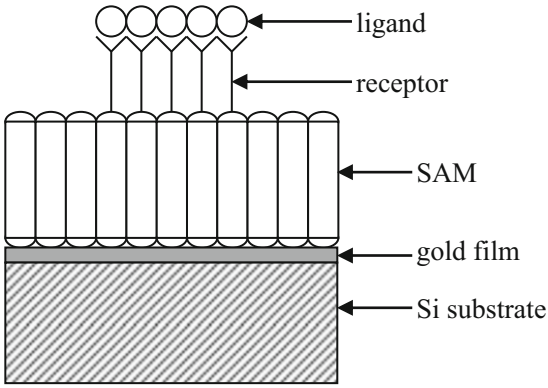


Fig. 10.3 Schematic of SAM enabled molecular recognition

The SAMs of thiolate compounds have many attractive features and offer several advantages over other SAM systems. They provide the highest structural order by densely packing long alkanethiolate chains on the surface and forming crystalline or quasicrystalline monolayers. They are easy to prepare, have density of defects low enough to be useful in many applications and good stability under ambient laboratory conditions, and, most importantly, are resistant to acid erosion. For example, silver surfaces with octadecanethiolate monolayers could be kept in ambient conditions without tarnishing for many months, and copper surfaces coated with the same monolayer sustain nitric acid attack. Gold surface protected by a monolayer of hexadecanethiol (HDT) is 10^6 times slower in cyanide (gold etchant). Such a property makes the SAM film an excellent ultrathin resist or passivation layer for pattern transfer with nanometer dimension. SAMs are also very useful in surface engineering, i.e., modification of surface properties such as wetting, adhesion, lubrication, corrosion, nucleation, protein adsorption, and cell attachment.

Once the gold surface is covered with organosulfur SAM which consists of long chain polymer molecules, other organic molecules can be added to it. So the surface becomes functionalized. The SAM surface can be functionalized in such way that it has receptor-terminated alkanethiolate groups which only recognize certain other molecules, as schematically illustrated in Fig. 10.3, forming a molecular recognition system [11]. Specific recognition systems can be defined by different interfacial binding property of molecules. It is like a "lock and key" system. One receptor only recognizes a specific group of molecules [12], which is a very useful technique to construct sensors [13].

10.3.2 Self-Assembled Supermolecules

Not only the SAMs, the power of chemical synthesis also provides access to a potentially vast range of functionally and structurally diverse molecular building

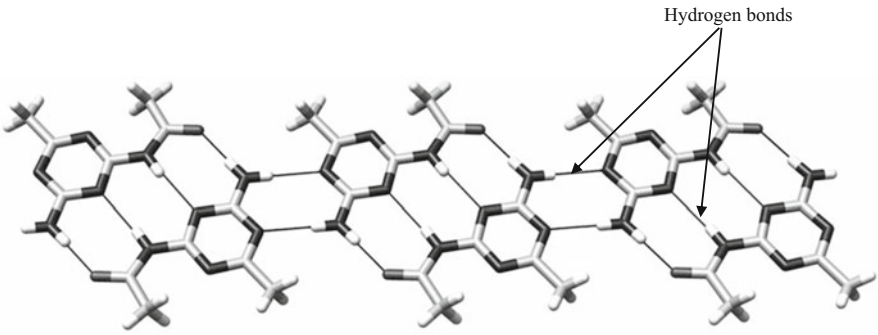


Fig. 10.4 Supramolecular structure self-assembled by hydrogen bonds

blocks, which can be linked through different types of relatively weak, non-covalent interactions (predominantly hydrogen bonds and metal-ligand interactions) to yield organized supramolecular architectures with tailor-made properties [14]. Figure 10.4 shows how groups of molecules are linked by hydrogen bonds. There are three groups of molecules, each of which is linked by two hydrogen bonds. Within each group, two similar molecular structures are linked by four hydrogen bonds. The hydrogen bond is a special chemical bond which is stronger than van der Waals force but much weaker than covalent bonds (1/20 of covalent bond strength) or ionic bonds [15]. Because of the low energy required for hydrogen bonding to occur, molecules can be linked to assemble from simple molecular structures to supramolecular structures at normal conditions, provided those molecules are mobile. Supramolecular engineering has enabled endless possibilities to create complex nanostructures from simple molecular building blocks [16], which has found applications in a variety of fields, such as in high-performance supercapacitors [17].

Although hydrogen bonding is able to form supramolecular crystals or supramolecular compounds in solution, which relies on molecular mobility in solution phase and chemistry of intermolecular non-covalent bond, nanofabrication is more about constructing structures on a substrate surface. To construct supramolecular structure on a surface, the influences of the substrate atomic lattice and substrate electronic or chemical properties on non-covalent bonds also play critical roles [18]. The parameters crucial for this type of self-assembly are the surface mobility of molecules, their lateral interactions, and their coupling to the surface atomic lattice. Though there are more constrains in assembly of molecules on a solid surface than in a liquid solution, it provides the possibility to tune the surface property and to produce tailor-made molecular structures. One example is the construction of surface supported metal-organic coordination networks, resulting from the interaction of organic linker molecules and transition metals at surfaces with the coordination behavior of benzene polycarboxylic acid species and iron (Fe) atoms on a copper Cu(100) surface [19]. It involves sequential deposition of polyfunctional organic molecules and Fe atoms on an atomically clean Cu(100) substrate. Another example is the construction of single-atom transistor, as schematically shown in Fig. 10.5 [20]. It is

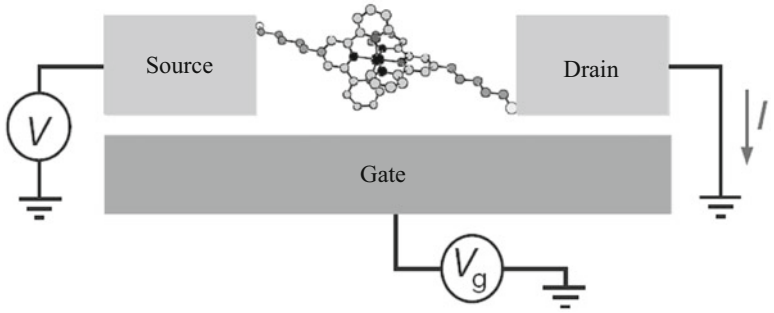


Fig. 10.5 Single-atom transistor constructed by self-assembly process

a supramolecular structure with a cobalt atom at the center of metal-organic coordination network. The cobalt atom becomes the quantum island which enables electron tunneling from source to drain electrodes in the transistor.

10.3.3 Evaporation Induced Self-Assembly (EISA)

While the SAM process produces a uniform dense layer, the same solution-phase self-assembly can also create mesoporous layer with highly regular nanostructures via evaporation induced self-assembly (EISA) method [21]. The EISA process involves the natural and spontaneous ability of surfactants to organize solutes into different shapes due to interactions such as van der Waals and electrostatic forces, hydrogen bonding, and π-π interactions. In the formation of mesoporous thin film by the EISA, as illustrated in Fig. 10.6 [22], an initial homogeneous solution consisting of surfactant (structure agent), soluble metal alkoxide or metal salt (inorganic precursor), and solvent such as alcohol or ethanol, water, and often acid (usually HCl) are mixed thoroughly. The solution is cast on a substrate by spray, spin, or dip coating. Evaporation of the volatile components in the liquid film takes place at the interface between air and the film. The repulsion force between molecular units due to the surfactants keeps them in equal distance. Final removal of organic surfactant by heating at calcination temperatures of 40–550 °C leaves a latticelike porous inorganic network on the substrate surface [23].

The mesoporous structures made by EISA have found a variety of applications including as nanoreactors, as adsorbents for removing volatile organic compounds, for removal of toxic metal ions, as organic dyes, and as bulky proteins. In addition to applications as catalysts and adsorbents, silica nanoparticles with ordered mesopores have attracted considerable attention as nanometer-sized containers for intracellular drug delivery; ordered mesoporous carbon materials have been investigated as metal-free catalysts for oxygen reduction reactions; mesoporous transition metal oxides such as titania and tin oxide have been used as electrodes in dye-sensitized solar cells, lithium-ion batteries, gas sensors, and supercapacitors [24].

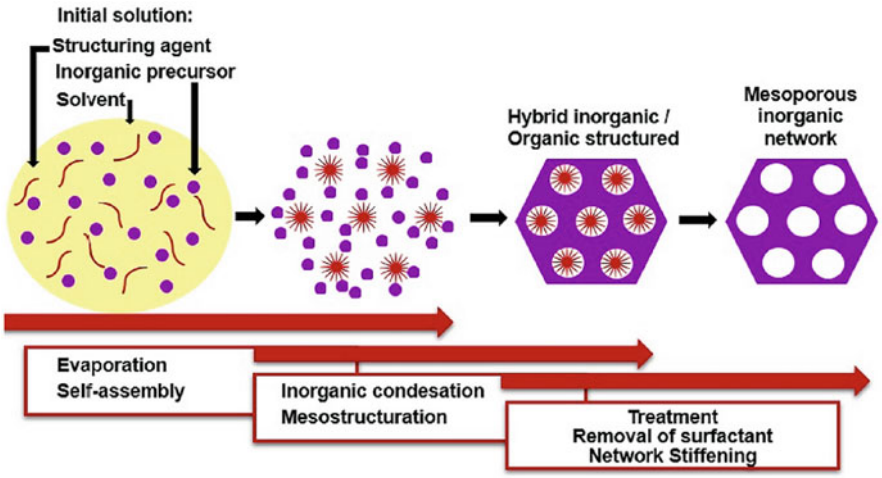


Fig. 10.6 The most representative steps of evaporation induced self-assembly (EISA). (Reprinted from [22] with gratitude to the authors and InTech)

10.3.4 Self-Assembly of Metal-Organic Frameworks (MOFs)

Another family of organic-inorganic hybrid porous nanomaterials created by solution-phase self-assembly of molecules is the metal-organic framework (MOF), which represents a revolution in chemistry due to the immense chemical and structural possibilities this family of materials can offer [25]. MOFs, also called porous coordination polymers, are self-assembled structures composed of metal nodes (metal ions or clusters) and polydentate organic ligands (mainly carboxylates, imidazolates, or phosphonates). The synthetic process generally consists of mixing the molecular building blocks, i.e., precursors, in a solvent (commonly methanol, dimethylformamide, or water). Then the mixture is either left undisturbed at room temperature for a period of time when the self-assembly process takes place or heated up by conventional heating or through other energy sources, such as microwave or ultrasound, at an appropriate temperature for a specific time. The incorporation of a modulator agent (i.e., additive) into the synthetic mixture can achieve a fine control over the self-assembly process to ensure the physical properties such as crystallinity, particle size, morphology, porosity, defectivity, and surface chemistry of MOFs. The topologies, porosities, functionalities, and surface properties of MOFs can also be manipulated by playing with different metal nodes and organic linkers and by means of post-synthetic modifications [26]. The resulting ordered structural motifs, similar to cage-like structures with pore size of 2–50 nm, provide MOFs with a high internal surface area and porosity, which are the key features exploited in the applications as diverse as gas storage and separation, molecular sieving, energy storage and conversion, sensing, catalysis, and controlled delivery of active species (e.g., drugs or fertilizers) [27].

10.4 Self-Assembly of Colloidal Particles

In addition to molecular self-assembly, another broad category of self-assembly is the formation of lattice-like structures by self-assembled colloidal particles or spheres. The self-assembled structures are often called colloidal crystals [28]. Colloidal self-assembly refers to self-assembly of particles or spheres with diameter from micrometers to nanometers in a liquid suspension. Particles in such a dimensional range have properties neither of individual atoms nor of bulk solid-state materials. There are three conditions which have to be satisfied to realize self-assembly of colloidal particles.

1. The colloidal particles must be able to move freely. These particles are normally dispersed in a liquid to form a colloidal suspension. That is how the name of colloidal particles came.
2. The particles must be sufficiently small so that their movement in a liquid is a Brownian motion. The Brownian motion enables the particles to interact sufficiently, which is the basis for self-assembly to happen. Normally the size of particles should be less than 5 μm. If the particles are too big, gravitation force may come to play, which will disrupt the Brownian motion of particles.
3. The shape of colloidal particles in the self-assembly process. It influences (i) the geometry of their close-packing, (ii) the directionality of the interactions, (iii) the selectivity of the interactions, (iv) the connectivity between objects in their assembly, (v) the attraction between them, and (vi) the dynamic aspects of their assembly, especially when the interaction is mediated by a fluid [29]. Generally, the particles should be spherically shaped with uniform diameter. For nonspherical particles, it is possible to control the outcome of self-assembled structures via tuning of intermolecular interactions through a meticulous design of particle shape and other environmental conditions [30].

Self-assembly of colloidal particles is a physical process, and there are several different forces which can promote the formation of colloidal crystals. The forces include capillary force or surface tension force which drive the particles to aggregate. Electrostatic force due to electric charge carried by particles and van der Waals force will also make the particles attract to each other. One of the perfect examples of colloidal self-assembly by capillary force and surface tension force is the formation of “coffee ring” [31]. When a drop of coffee liquid dries on a solid surface, it leaves a ringlike stain which consists of coffee particles. Similarly, micron sized polystyrene spheres in liquid suspension can form 2D ordered aggregation when the liquid layer is drying on a solid surface. The process is illustrated in Fig. 10.7. Initially they randomly scatter in the liquid layer. As the liquid evaporates, the polystyrene spheres are pushed to the edge of liquid drop whose boundary is “pinned” to the surface, which means that when the liquid evaporates, the liquid drop cannot shrink in circumference but instead flattens out. That flattening motion, driven by surface tension, pushes liquid flowing to the edge and carrying anything suspended in it, such as the polystyrene spheres or coffee particles, to its edges, as illustrated in

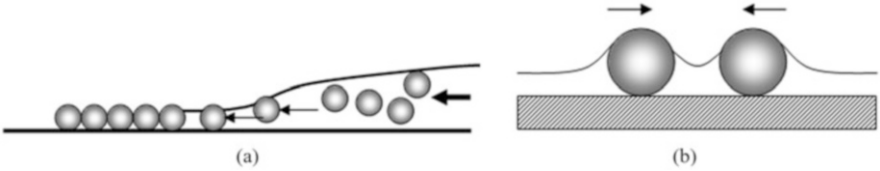


Fig. 10.7 Schematic of self-assembly process of "coffee ring": (**a**) surface tension driven aggregation, (**b**) capillary force driven aggregation

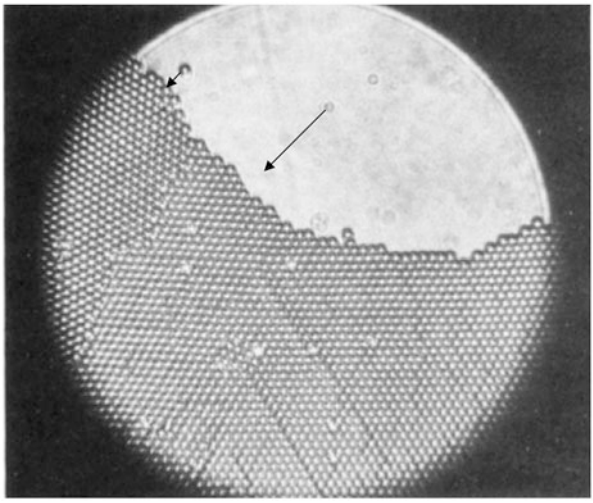


Fig. 10.8 Orderly packed polystyrene spheres formed by surface tension and capillary force driven self-assembly. (Reprinted from [32] with permission)

Fig. 10.7a. Meanwhile, the spheres or particles coming closer to each other get a further push by capillary force when liquid surface lowers below the diameter of particles, as illustrated in Fig. 10.7b. By the time the liquid fully evaporates, which may take hours, most of the particles have reached the edge and closely packed. Figure 10.8 shows the optical image of orderly packed polystyrene spheres assembled from its initially random distribution. The ordering is due to surface tension of suspension liquid and capillary force between the particles, which pulls the particles together.

The "coffee ring" type of self-assembly has been used to create colloidal crystals in large area. The basic principle is the same, coating colloidal particle suspension liquid onto a surface and letting it dry. The coating can be done by spinning, inkjet printing, dip coating, or electrospraying [33]. As aforementioned, the self-assembly of colloidal particles only happens when liquid surface drops below the diameter of particles; controlling the drying process is the key to achieve high quality self-assembled colloidal crystals. In addition, the concentration of particles in the suspension liquid is also important for self-assembly of 2D colloidal crystal. Too low the concentration will result in many defects in the assembled lattice. Too high the concentration will create locally multilayer pilling of particles. Figure 10.9 shows the

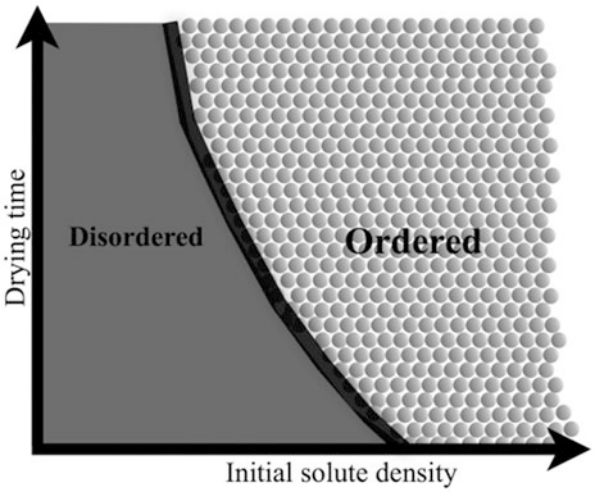


Fig. 10.9 Proper drying time and particle concentration (solute density) to produce ordered self-assembly [34]

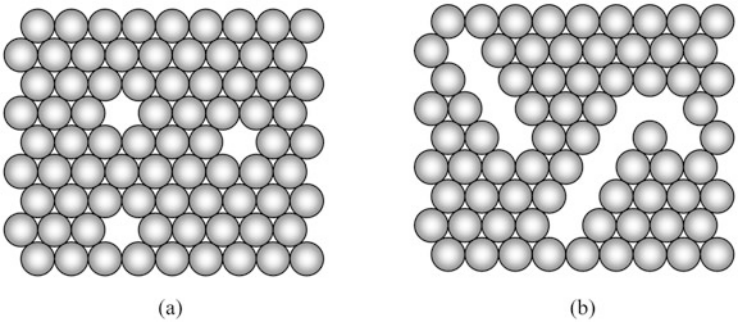


Fig. 10.10 Illustration of defects in self-assembly of colloidal crystal: (**a**) microscopic defects (vacancies), (**b**) macroscopic defects (dislocations)

relation between drying time and concentration of colloidal particles [34]. There is a boundary of dry time and solute density, only within which an ordered self-assembly can be achieved. The same process for 2D assembly of colloidal crystal can be extended to produce 3D colloidal crystal. It is simply a repeating of 2D assembly layer upon layer to realize 3D assembly. In addition to the drying induced self-assembly, the assembly can also take place with the assistance of Marangoni flow which is the phenomenon of liquid flow from regions of low surface tension to those of high surface tension [35].

Self-assembly of colloidal crystals is not a totally controllable process. There are inevitably defects in the assembled structures. The defects can be classified into two types, microscopic and macroscopic defects, as illustrated in Fig. 10.10. Microscopic defects are the missing individual particles in the assembled crystal lattice, called vacancies, as seen in Fig. 10.10a. Macroscopic defects are the cracks appearing in

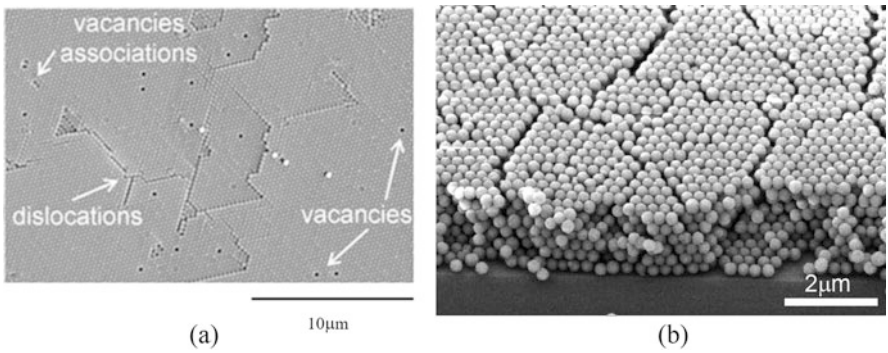


Fig. 10.11 (**a**) Defects in 2D colloidal crystal, (**b**) defects in 3D colloidal crystal

the bulk of crystal, called dislocation, as seen in Fig. 10.10b. Examples of such defects in self-assembled colloidal crystal are shown in Fig. 10.11 where Fig. 10.11a is a 2D colloidal crystal with vacancy and dislocation defects [36], and Fig. 10.11b is a defected 3D colloidal crystal [37]. It is a fact that there are always defects in self-assembled colloidal crystals. A high quality self-assembled colloidal crystal may have low level of defects. One can only control the defect density to its minimum, but cannot completely eliminate them.

10.5 Guided Self-Assembly

Pure self-assembly or bottom-up approach cannot generate many useful nanostructures, apart from forming latticelike periodical arrays, as shown in Fig. 10.11, which may only have practical applications as photonic crystals [37, 38] or in magnetic storage [39, 40]. As described in the previous sections, self-assembly is strongly influenced by external energy and environmental conditions, as those snowflakes shown in Fig. 10.1. By manipulating these conditions and external energy, one can direct the assembly to develop toward a designed path. Or in other words, the self-assembly process can be guided.

10.5.1 Surface Topography

The obvious guided self-assembly is to use a topographical template. Conventional top-down fabrication can provide a template with patterned structures. The fabricated topographical structures create physical boundaries to prevent self-assembly from spreading all over the surface. Patterned surface topography can be quite effective to trap nanoparticles. Figure 10.12a illustrates schematically how

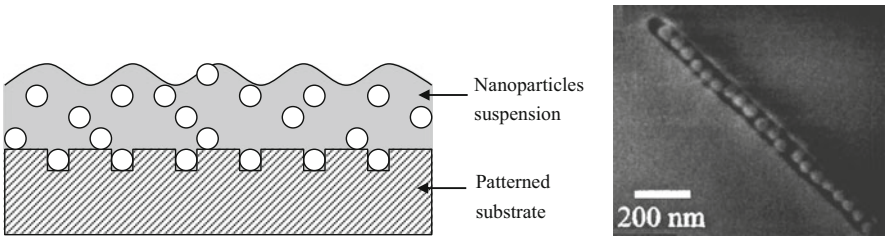


Fig. 10.12 Guided self-assembly of nanoparticles by surface topography: (**a**) schematic illustration; (**b**) SEM image of assembled nanoparticles in a patterned groove. (Reprinted from [42] with permission)

nanoparticles are guided into pre-patterned areas when the substrate is immersed into a nanoparticle suspension. This is only a simplistic illustration. The real process of trapping the nanoparticles is much complicated. These nanoparticles are not physically falling into the traps, because the gravitation force is too weak for these nanosized particles, but driven by capillary force of the suspension liquid. It only happens when the suspension liquid drop dries up by evaporation, as illustrated in Fig. 10.7, and the surface topography pins the retreating liquid surface as it shrinks, bringing the nanoparticles into the trap [41]. Figure 10.12b shows the SEM image of self-assembled nanoparticles in a groove [42]. Once they aggregate in the trap, electrostatic force, if the suspension is ionic solution, or interparticle van der Waals force will further secure the assembly and their attachment to the substrate before the drying completed. The assembly process can be sped up by heat induced interfacial convective flow [43], which reduces the assembly time to a few minutes as comparing to leaving the suspension liquid drying up at ambient condition, which can take hours.

Alternatively, nanoparticle suspension can also be coated onto a topographical substrate by dip coating, i.e., the substrate is dipped into the suspension and then slowly pulled out, leaving a thin layer of liquid containing nanoparticles on the surface, like the formation of Langmuir-Blodgett film. In this way, uniform assembly of nanoparticles into patterned grooves can be achieved by balancing the pulling speed and evaporation rate of the thin liquid film [44]. The assembled structures are closely related to the geometry of topographical structures. The geometry of grooves must be much larger than the diameter of colloidal particles [45]. If their dimensions are comparable, then the geometry of topographical structure made by top-down nanofabrication must match the integers of particle diameter [46]. The final assembled structures are very much dependent on the top-down fabricated topographical structures [47]. Not only 2D array of colloidal particles can be guided by top-down fabricated holes, but more layers can be assembled on top of the initial 2D array. For example, up to four layers of 500-nm-diameter polystyrene (PS) nanospheres were assembled on patterned photoresist holes by controlling the concentration of PS nanospheres in the aqueous solution [48].

Surface topography can also be created by bubbles whose ridges define the boundaries of surface pattern structure. Silver nanowires were dispersed in water with a surfactant and a thickening agent. By stirring the mixture to generate a large number of bubbles, the silver nanowires were lining up along the bubble ridges. In this way, a transparent conductive layer was produced [49]. The bubbles can be further regulated by top-down fabricated surface topography so that the lineup of nanoparticles through self-assembly along the bubble ridges forms predesigned patterns [50].

10.5.2 Surface Energy

In addition to topographical template, the substrate surface can be physically or chemically modified so that colloidal particles can be selectively assembled. This can be achieved by changing the surface energy, i.e., the surface hydrophobicity. In this way, nanoparticles in a colloidal suspension will only assemble onto the hydrophilic surface area, not stick to the hydrophobic surface area.

Surface hydrophobicity and hydrophilicity, or another name is wettability, is a very important branch of surface science. The wettability of a surface is defined by water contact angle. A surface is hydrophobic when its static water contact angle θ is $>90°$and is hydrophilic when θ is $<90°$ [51]. There are many ways to modify the wettability of a solid surface. Physical modifications include making a surface rough or porous to increase its hydrophilicity. Chemical modifications are mainly adding chemically active molecules to increase its surface energy. For example, silicon surface is naturally hydrophilic because a thin native oxide forms over time in air, and it becomes hydrophobic by removing the native oxide in hydrofluoric acid (HF). To guide the self-assembled colloidal particles into a pattern, the key parameter is the contrast between hydrophobicity and hydrophilicity of adjacent surfaces, which determines the contrast of self-assembled patterns [52]. One way to achieve the high contrast is to introduce a patterned hydrophobic self-assembled alkane phosphate monolayer (SAM) on SiO_2 area. The pattern created with the hydrophobic SAM has the water contact angle of 110°, while the hydrophilic SiO_2 area has contact angle of only $<6°$ [53]. There are other ways to chemically pattern a surface such as by microcontact printing (μCP) which has been discussed in length in Chap. 6 (Sect. 6.6.2) or UV illumination through a mask to induce modification of hydrophobic surface into hydrophilic surface [54].

10.5.3 Electrostatic and Magnetic Forces

At molecular level, electrostatic self-assembly is based on the attraction of oppositely charged ions where the interactions can be between cationic and anionic molecules or between zwitterionic molecules which comprise positively and

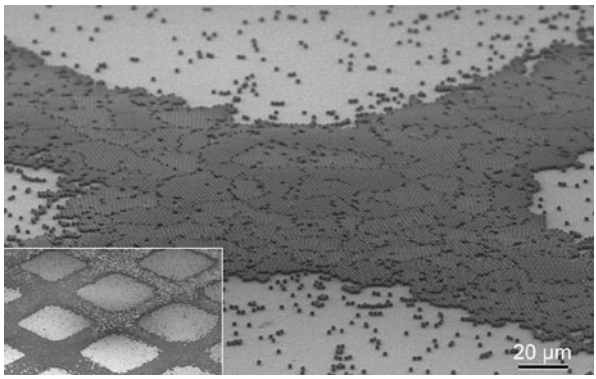


Fig. 10.13 Two-dimensional pattern formed by self-assembled polystyrene spheres; the inset image shows the overall appearance of the pattern (~600 μm). (Reprinted from [61] with permission)

negatively charged functional groups that are covalently bonded with each other. Two oppositely charged particles suspended in a fluid will attract to each other. In conjunction with other aforementioned forces, such interactions only result in formation of microscopic structures such as supermolecules [55], rather than meaningful macroscopic patterns.

On the other hand, charged particle in a suspension fluid will attract to electrodes which is oppositely charged. This is a long range force compared to aforementioned molecular attraction. Substrate with uniform surface charge can be made by immersing clean gold surfaces in ethanol solution of alkanethiols for 10 min. The surface can be either positively charged or negatively charged, depending on the type of thiol groups in the solution. Patterned surface charge can be made by μCP with different thiol compounds as the ink [56] or contact electrification [57]. Polystyrene sphere (PS), which is the mostly used particles for controlled self-assembly, can also be made to carry electric charge. Carboxyl-terminated PS can be made to carry negative charge, and amidine-terminated PS can be made to carry positive charge.

For electric force assisted self-assembly, the oppositely charged nanoparticles are attracted to the charged substrate surface in the same way as electrophoretic deposition. However, they may not assemble into orderly packed colloidal crystalline structure, because these particles repel each other due to the same type of electric charge they carry. With an externally controllable electrical field, an orderly packing can be formed, for example, by raising the applied voltage [58]. The exact nature of the assembly is not clear but is believed due to some electrohydrodynamic [59] or electroosmotic [60] effect near the electrode surface at high field. The behavior of charged nanoparticles can also be altered by UV illumination of a transparent electrode surface, which also helps the formation of colloidal crystal structures. Figure 10.13 is a SEM image of assembled polystyrene spheres (PS) by UV assisted electrophoretic deposition. The assembly is not perfect as there are only localized 2D crystalline lattice (ordered) structures within the patterned area. Each of the crystalline domain is around 10–20 μm. However, the overall goal of selective assembly of PS onto pre-patterned area has been achieved, as seen from the inset image.

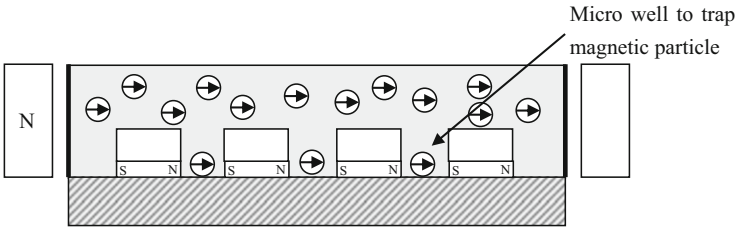


Fig. 10.14 Schematic of magnetically assisted self-assembly of magnetic particles

Similar to electrostatic force, the magnetic force is also long range force. Magnetic particles can be assembled onto a magnetized surface. Unlike electric field, magnetic field has practically no effect on most molecules or chemical processes. In addition, while materials may become polarized by electric field, magnetic field can permeate uninhibited through most mediums, particularly through liquids and gases. Magnetically assisted self-assembly mainly refers to assembly of superparamagnetic colloidal particles whose magnetization is linearly related to the applied magnetic field, i.e., no hysteresis exhibits in this type of materials. Superparamagnetic colloidal particles remain stably suspended in the absence of magnetic fields because they are unmagnetized and thus experience virtually no magnetic interactions with other particles. Particles made of ferromagnetic materials becomes superparamagnetic when their size is smaller than the superparamagnetic critical size, which is roughly 10 nm in diameter [62]. Superparamagnetic colloidal particles are typically 100 nm to 10 μm in diameters, composed of superparamagnetic nanoparticles encapsulated or dispersed inside silica or various polymers, such as polystyrene or polyvinyldibenzene. Increasing the magnetic content of the particles by encapsulating more magnetic nanoparticles inside each polymer bead results in only an increased magnetic response without altering its superparamagnetic properties.

Magnetic force as an external power source has been exploited for magnetically induced self-assembly. The energy transmitted by an external magnetic field to magnetic nanoparticles leads to the changes of orientation, migration, and aggregation to form microscopically ordered colloidal crystals, which is much rapid and reversible [63], compared to the conventional surface tension and capillary force induced self-assembly introduced in Sect. 10.5. For the formation of macroscopic pattern by superparamagnetic colloidal particle assembly, combined actions of external and local magnetic fields are required. As illustrated in Fig. 10.14, the particles suspended in a liquid are first magnetized by an external magnetic field. These particles then become magnetic dipoles aligning to the external magnetic field lines. In the meantime, they are also subject to the local magnetic force generated by patterned magnetic elements on the template, which either attracts or repels these magnetic dipoles. These magnetic particles are then selectively assembled into the template patterns or trapped by the local magnets. Since each magnetic element on the template can be individually magnetized, just as the data recording on a magnetic disk, the assembling of magnetic particles can be programmed to form nonperiodic patterns [64].

10.5.4 DNA Templates

Deoxyribonucleic acid (DNA) is a long chain of polymer made from repeating units called nucleotides. Biologically, DNA contains the genetic instructions used in the development and functioning of all known living organisms. DNA becomes an attractive building block in nanofabrication because of its unique physical and chemical properties. It was long ago discovered that symmetrical DNA patterns such as squares and stars can be constructed by repeatedly folding the long natural DNA single strand and fixing it with the short strand like origami [65]. The principal mechanism relies on the ability of DNA strands to cross over from one double-strand DNA helix to connect to another adjacent helix. These "crossovers" can be programmed to occur at specific locations using sequence design and base pairing of single-strand DNA [66]. The so-called DNA origami can be used to construct the most complex DNA nanostructures, which subsequently serve as a 2D template or 3D scaffolds to assemble functional nanomaterials or molecules, such as metallic nanoparticles and nanowires and carbon nanotubes. DNA-templated nanofabrication is a rapidly developing discipline that produces custom-designed nanomaterials through the use of DNA molecules as self-assembly guides [67]. DNA guided assembly of conductive and semiconductive nanoscale entities has been particularly useful in constructing nanoelectronic devices [68] and 3D ordered inorganic frameworks [69].

10.6 Nanosphere Lithography

Two of the most studied colloidal particles for self-assembly are polystyrene (PS) and silica nanospheres. Apart from the assembled colloidal crystals being used as photonic crystals [70], they also can serve as masks for the so-called nanosphere lithography [71]. As illustrated in Fig. 10.15, self-assembly of PS spheres with uniform diameter leaves gaps (Fig. 10.15a, b). Materials deposited by thermal evaporation or plasma sputtering over the spheres-covered surface can only reach the substrate through these gaps (Fig. 10.15c). After removing the polystyrene spheres in CH_2Cl_2 acid, similar to the liftoff process introduced in Chap. 8, the substrate surface is left with array of triangular dots (Fig. 10.15d). Assuming the sphere diameter is D, the dot size formed through a single layer PS sphere mask is 0.233D, and the distance between dots is 0.577D. A double layer of spheres can further reduce the gap [72], hence the size of dots which are calculated to be 0.155D. Instead of deposition in normal direction, which only produces dots, deposition in glancing angles can form surprisingly complicated patterns in the so-called shadow sphere lithography [73]. In addition to making dot array by material deposition, reactive ion etching can produce hole array through the nanosphere mask [74]. However, both the dot/hole size and the period of dot/hole array are restricted by the available size of polystyrene spheres which are normally above 200 nm for reliable assembly [75]. The lack of flexibility is the major drawback of this technique.

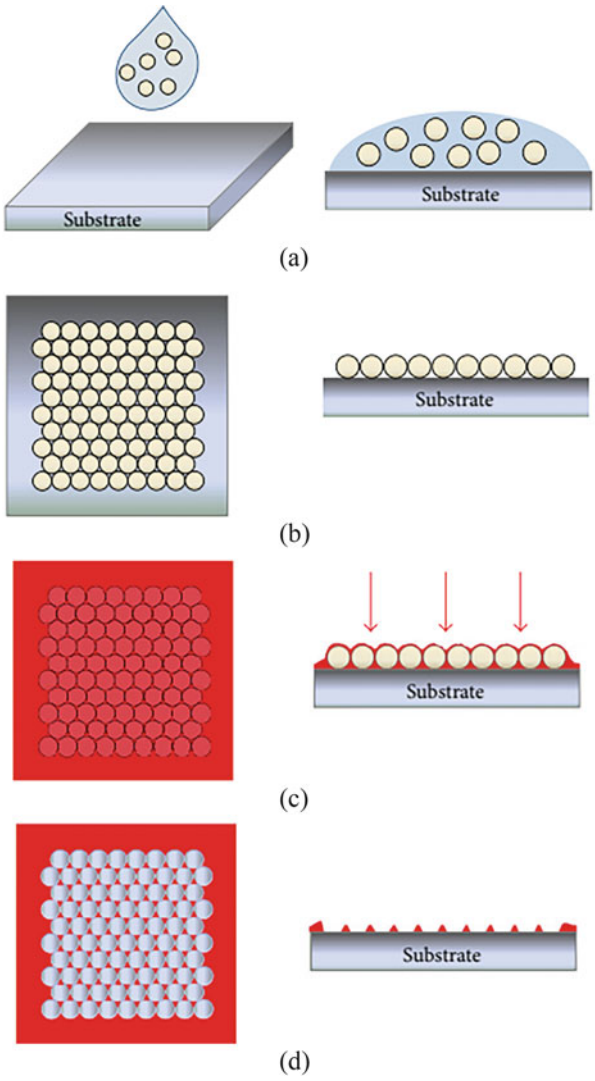


Fig. 10.15 Nanosphere lithography steps: (**a**) suspension liquid of PS spheres casting on a flat substrate surface; (**b**) self-assembly of PS spheres into 2D hexagonal-close-packed array; (**c**) deposition of a functional material through the gaps; (**d**) liftoff of PS spheres [71]

Another kind of nanosphere lithography utilizes self-assembled polystyrene or silica nanospheres on a substrate surface as a lens array. By shining UV light onto the transparent nanosphere lens array, light spots are formed under each nanosphere, and a regular array of holes can be created if the underlayer is photoresist [76]. Figure 10.16 illustrates the process of nanosphere optical lithography. With illumination wavelength of 400 nm and conventional photoresist AZ5214, holes with minimum diameter of about 250 nm were formed in the photoresist layer using silica sphere of ~0.97 μm diameter, which was about 0.6 times of the exposure wavelength. In

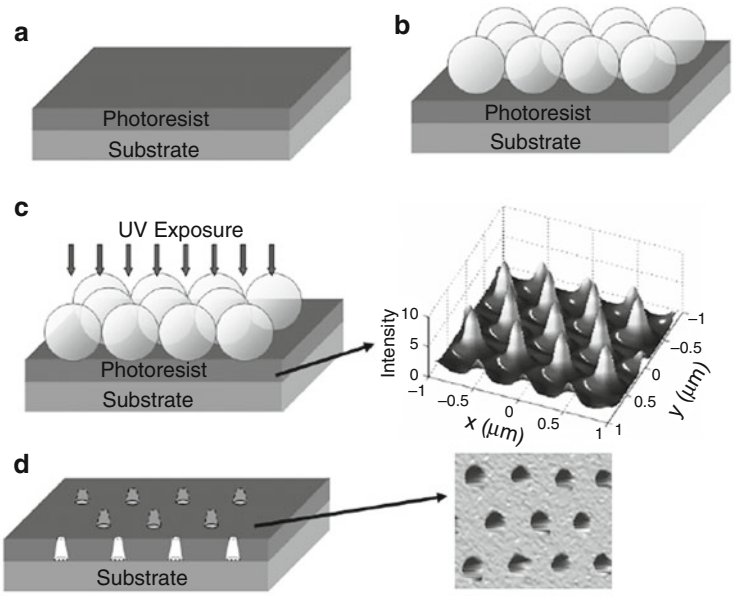


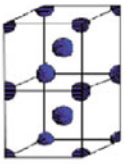
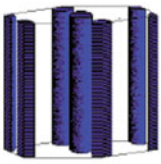
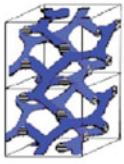
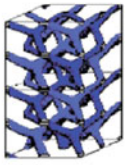
Fig. 10.16 Schematic of nanosphere optical lithography. (Reproduced from [76] with permission)

addition to normal exposure, tilted exposure produced various shapes of feature in photoresist, and exposure in thick resist layer resulted in 3D structures due to the Talbot effect in Z direction [77].

10.7 Block Copolymer Lithography

Block copolymers (BCPs) consist of chemically dissimilar polymer chains covalently linked to form a single molecule [78]. Under certain conditions, the dissimilar blocks tend to segregate into different spatial domains, called microphase separation [79], similar to oil-water separation. On the other hand, the spatial separation is constrained by the chemical connectivity of the blocks. As a result of these competing effects, self-organized periodic structures are created on a nanometer scale. Formation of any particular microdomain structures depends on relative volume ratio between blocks and chain architecture as well as the persistence lengths of the respective blocks. In the simplest case of noncrystalline flexible coil AB diblock copolymers, the composition of the AB diblock (i.e., the volume fraction of block A and B) controls the geometry of the microdomain structure. Table 10.1 gives some possible spatial microdomain structures due to different volume fractions of block A in the diblock copolymer. More complex spatial microdomain structures can be formed with triblock copolymers. The nanoscale spatial structures that can be formed from block copolymers are almost endless.

Table 10.1 Possible spatial microdomain structures for diblock copolymer with different volume fractions of one of the blocks

Nature of patterns	Spheres (SPH) (3D)	Cylinders (CYL) (2D)	Double gyroid (DG) (3D)	Double diamond (DD) (3D)	Lamellae (LAM) (1D)
Space group	$Im\bar{3}m$	p6mm	$Ia\bar{3}d$	$Pn\bar{3}m$	pm
Blue domains: A block					
Volume fraction of A block	0-21%	21-33%	33-37%		37-50%

Reprinted from [80] with permission

The relevance of BCP to nanofabrication is its ability to form nanoscale ordered pattern structures in a very simple and low cost manner. They are polymers same as photoresists. The regular array of microdomains is equivalent to the patterns made by photolithography; therefore, it is a different kind of lithography, i.e., block copolymer lithography. After nearly 30 years of development, BCPs are now able to create polymer patterns with sub-10 nm resolution [81].

10.7.1 Directed Microphase Separation

Block copolymers, if allowed to self-assemble alone, will form some sort of ordered structures due to the microphase separation from aforementioned competing effects. However, the microdomains are formed randomly and grow as a polygranular texture with different blocks and having sizes of several tens of nanometers, and the periodic ordering maintained only over distances of several tens of lattice constant (i.e., a typical grain size of only 1–2 μm). This short range ordering is of no interests to nanofabrication. To achieve long range ordered microdomain separation, both internal and external influences on the formation process should be applied. The internal control includes the choice of molecular weight, composition, and chain architecture. Table 10.1 has demonstrated how the microdomain structures vary simply due to the content of one of the blocks. The external control can be many

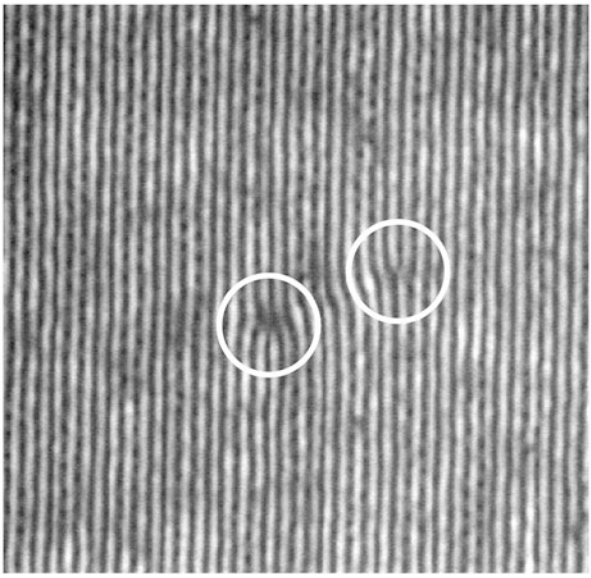


Fig. 10.17 Dislocation defects in BCP self-assembly of line and space pattern

ways. The most commonly used external influence is thermal annealing. At high temperature, the mobility of block polymer chains is enhanced, which facilitate the interaction between different blocks. For example, a polystyrene/polybutadiene (PS/PB) diblock layer became ordered spherical holes after annealing at 110°C for 1–12 h [82]. Other external influences include mechanical flow field [83] and electric field [84] or by the influence of surface interactions [85], chemically patterned surface [86], or temperature gradient to create a preferred orientation or alignment of microdomains [80].

Like all self-assembly processes, self-assembly of BCP will inevitably have defects. An example of defects is shown in Fig. 10.17 [87]. They are dislocation defects in a line-space microphase separation. One strategy to avoid defect formation in the course of self-assembly and to facilitate their removal is to direct the self-assembly by external fields such as aforementioned electric fields or shear flow, which are however only effective to some extent. As the probability of generating defects is proportional to the area the assembly takes place, an effective approach to having the defect level under control is to confine the assembly in a small area defined by chemical or topographical boundaries. These two techniques combine the advantages of traditional lithography to fabricate the chemical or topographical guiding pattern with the copolymer's ability to spontaneously form dense structures within the boundaries [88].

Self-assembly of BCP confined within the boundaries defined by surface chemical property is called chemoepitaxy, as illustrated in Fig. 10.18a, which involves the creation of dense chemical patterns on a neutral substrate to generate preferential wetting sites for one of the polymer ingredients in the BCP. Many methods can be used to selectively tune the surface free energy of a neutral surface, such as

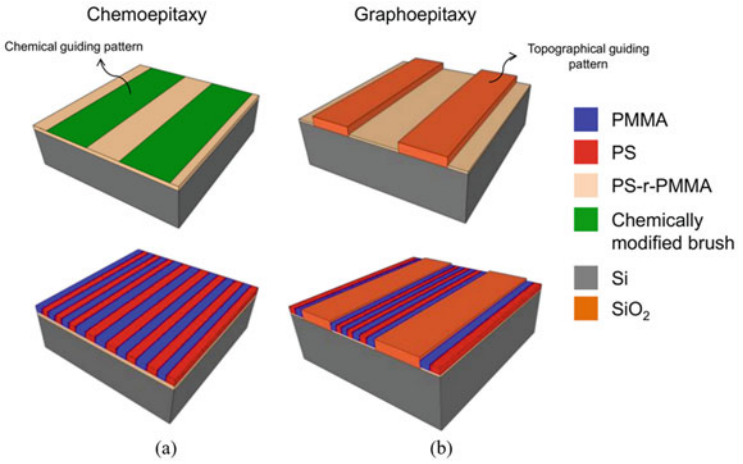


Fig. 10.18 Schematics of (**a**) chemoepitaxy and (**b**) graphoepitaxy for guided BCP assembly [90]

photolithography, electron-beam lithography, masked oxygen plasma functionalization, scanning probe lithography, and those introduced in Sect. 10.5.2.

Self-assembly of BCP confined within the boundaries defined by surface topography is called graphoepitaxy, as illustrated in Fig. 10.18b, which involves conventional top-down lithography to make low density patterns and having the assembly only in the zones divided by these patterns. An example of graphoepitaxy is shown in Fig. 10.19 [89]. Figure 10.19a is the line pattern by 193 nm photolithography. By coating PS-PMMA block copolymer over the patterned surface and annealing the BCP film, self-assembly took place between the patterned lines and added two more lines in between. The original pattern density was tripled, as seen in Fig. 10.19b.

10.7.2 Directed Self-Assembly for IC Manufacturing

The reason that self-assembly of BCP has been continuously studied for 30 years is its potential to be employed as a lithography technique for IC manufacturing. In IC manufacturing, interconnecting lines and contact holes are the two most important structures requiring high resolution lithography, which also happens to be the most common patterns produced by BCP self-assembly. Directed self-assembly (DSA) of BCP offers a simple and low-cost solution to extend the photolithography resolution beyond its diffraction limit. As demonstrated in Fig. 10.19, pattern density is increased with the same effect as that of multiple patterning of optical lithography (Chap. 2). In 2007, DSA was for the first time written in the International Technology Roadmap for Semiconductor (ITRS) as one of the alternative lithography techniques for IC manufacturing.

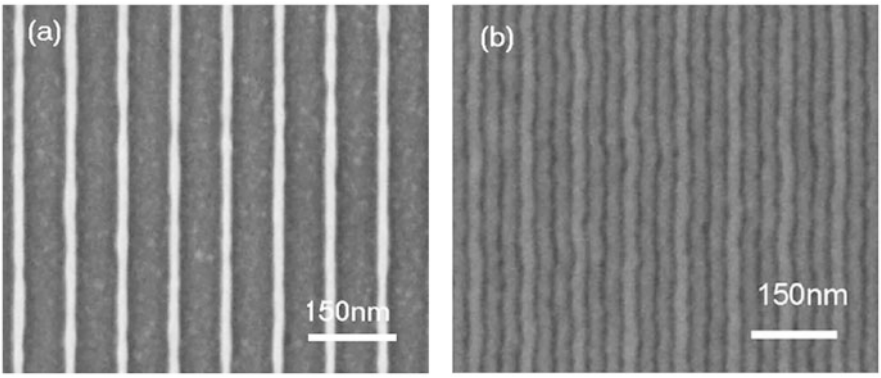


Fig. 10.19 Combination of (**a**) conventional photolithography and (**b**) BCP self-assembly to produce high density line and space pattern [89]

Among the many BCPs studied, polystyrene-block-poly-poly (methyl methacrylate) (PS-b-PMMA) is the most important BCP suitable for IC patterning, not only because it can generate both hole and line structures but also because the PMMA can be selectively removed after microphase separation which leaves the PS pattern as etch mask for subsequent pattern transfer [91]. Although there have been attempts to integrate BCP assembly to IC manufacturing on 300 mm wafer scale for 7 nm IC node [92], there are many practical issues remaining to be solved before the IC industry can safely use it. The issues include high level of defects, long annealing time, and thin assembled polymer layer [93].

Self-assembly of BCP is promoted by thermal annealing, which is a long process because the microphase separation involves the reorganization of polymer chains. The annealing process is normally in the range of a few hours to tens of hours. In the example of BCP self-assembly of line/space patterns shown in Fig. 10.19, the assembly took 2 days at 190 °C temperature in vacuum. This long process cannot be tolerated in IC manufacturing. To meet the throughput target of 120 wafers per hour, BCP assembly time is estimated to be no more than 6 min [94]. Various schemes have been developed to shorten the annealing time, including rapid annealing at elevated temperature [95] or solvent annealing [96]. Solvent annealing is based on a solvent vapor swelling the polymer and causing chains to move apart, thereby allowing fast separation. Currently, the cycle time is about 10 min but expected to get down to 1–2 min [97].

As the assembled PS-b-PMMA layer is very thin, only less than 10 nm in order to realize high resolution and rapid formation of self-assembled patterns, a hard masking layer is always used, which is in most cases SiO_2 or Si_3N_4. There are two steps of pattern transfer: selective removal of one polymer component (PMMA) and etching of SiO_2 or Si_3N_4 using the PS as patterned mask. The etching chemistry can be optimized to have high selectivity to keep the PS intact while only etching the PMMA with PMMA:PS etch selectivity >20 [98]. Once the PMMA is removed, the

PS pattern becomes the mask for second etching of underlying hard mask which is also very thin (~10 nm) [99].

The most critical issue which stops the DSA to be adopted in IC manufacturing is its defect level. To be relevant for industrial manufacturing, the total defect density should be below one defect per $1cm^2$ pattern area, which was very difficult to achieve. Primary contributors to the total defectivity are dislocations and bridges, formed during the self-assembly process [100]. Shortening the annealing time and confining the DSA in narrower boundary can all help to reduce the defectivity. For example, only 3x pattern density multiplication is allowed by chemoepitaxy or graphoepitaxy. There are also continued research efforts toward developing block polymers incorporating highly incompatible segments (termed "high χ") [81]. Currently, DSA can make patterns correspond to the 3 nm or 5 nm IC technology node. Combined with high-NA EUV, DSA may help to push the lithography further down to 2 nm node, as the 2022 edition of International Roadmap for Devices and Systems (IRDS™) envisioned [101].

10.8 Self-Regulated Porous Alumina

Anodization is an electrochemical process whereby certain metals (aluminum, niobium, tantalum, titanium, tungsten, zirconium) can be oxidized electrochemically. During anodization, the metal workpiece is the anode in an electrolyte bath. The cathode is any electrical conductor that does not react with the electrolyte. Once connected to a power supply, electrons attracted to the anode allow ions at the metal surface to react with water molecules in the electrolyte to form an oxide layer on the metal, and hydrogen gas is released from the cathode side. Normally, the electrolyte is selected in which the oxide is insoluble so that a barrier oxide is formed, which can serve as protection coating from corrosion or an electrical insulation layer. If the electrolyte is a diluted acid, the oxide layer will be partially dissolved and the oxidation surface becomes porous. This is how porous alumina or anodized aluminum oxide (AAO) is produced. Porous alumina possesses many desirable engineering properties such as excellent hardness and corrosion and abrasion resistance. The most visible application in everyday life is the durable coloring of aluminum surface in decorative and architectural applications [102].

Porous alumina became relevant to nanofabrication due to its ability to produce highly ordered pores with pore diameter of 10–400 nm and pore density of 10^8–10^{10}/cm^2. It is a self-regulated process, very much like the previously introduced self-assembly processes. Formation sequence of ordered porous alumina is schematically illustrated in Fig. 10.20. When the current starts to pass through the aluminum anode, a barrier layer of nonporous alumina begins to form (I). During anodization of aluminum in an acidic electrolyte, there are two competing processes taking place—one is the oxidation to form alumina layer, and another is the dissolution of alumina in acid. If there is no dissolution of alumina, the oxidation process will eventually stop because the surface becomes insulating as the oxide layer becomes

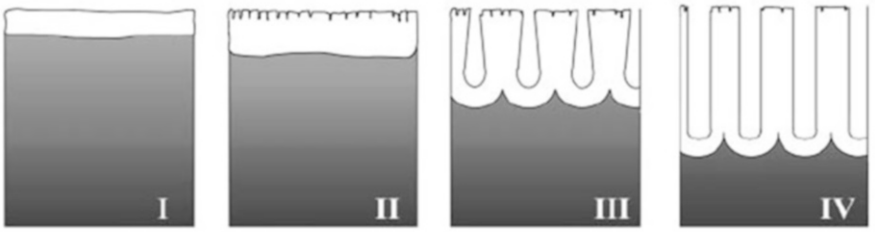


Fig. 10.20 Sequence of porous alumina formation. (Reprinted from [104] with permission)

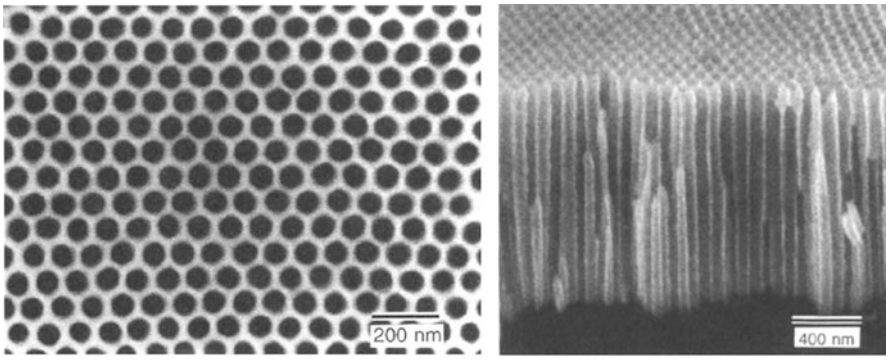


Fig. 10.21 Hexagonal nanopore array formed by aluminum anodization: (**a**) top view; (**b**) cross-sectional view. (Reprinted from [103] with permission)

thicker. As this layer grows thicker, the electrical field across the oxide layer increases due to the increase of resistance. The high electrical field causes small fractures on the top of the oxide layer (II). Nanopores are formed initially from these fractures in the alumina surface, where acid penetrates and dissolves alumina and some cracks develop further into large pores (III). Because the starting points of alumina dissolution are random, the nanopores initially formed are also random. However, as the oxidation and acid dissolution progress, a balance between the competition of oxidation and dissolution is reached, and ordered hexagonal structures start to emerge (IV). The formation of porous surface also decreases the resistance, which produces a voltage decrease when working at constant current density. Finally, a constant voltage is maintained due to the balance of oxidation-dissolution at the bottom of the pores. Figure 10.21a shows highly ordered hexagonal holes formed in alumina, and Fig. 10.21b is the cross-sectional view of the porous alumina, showing very dense and high aspect ratio honeycomb structures [103].

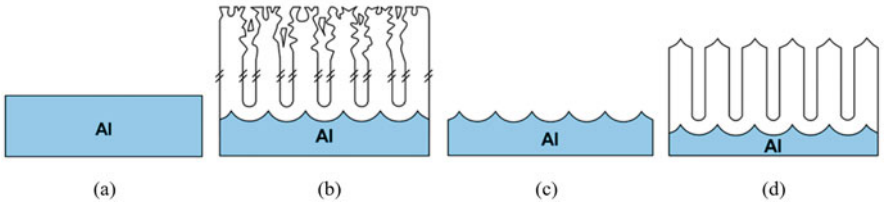


Fig. 10.22 Two-step anodization of aluminum for the formation of regular pores: (**a**) initial aluminum surface; (**b**) first anodization; (**c**) oxide removal; (**d**) second anodization. (Reprinted from [107] with permission)

The formation process described in Fig. 10.20 is too slow. For example, the nanopore array shown in Fig. 10.21 was formed after 160-h anodization of pure aluminum in 0.3 M oxalic acid and 40 V voltage at 0 °C temperature. High voltage and high electrolyte temperature, which results in high anodization current density, will cause the formation of nanopores faster. The growth rate of porous alumina as high as 10 μm/min was reported at voltage of 195 V and current density of 150 mA/cm^2 [105]. However, the pore size also grows bigger at higher current density. The best way to speed up the nanopore formation process while maintaining small pore size is to anodize in two steps [106], as schematically shown Fig. 10.22 [107]: the first step is to anodize the aluminum to generate disordered pores (Fig. 10.22a, b); then the porous top part is removed in acid which reveals the textured surface with arrays of almost hemispherical concaves (Fig. 10.22c); the regular surface texture is then used as the starting points for second step anodization which finally forms the regular pores (Fig. 10.22d) with much faster formation rate of 1–2 μm/min [108]. The two-step anodization can reduce the pore formation time to one-tenth of the one-step process. Many factors can affect the geometry of porous alumina structure such as pore diameter, pore spacing, and pore density, which include anodization potential, electrolyte type and concentration, stirring of electrolyte, anodization temperature, and time [109].

Since the nanopores are initiated from surface dents, as illustrated in Fig. 10.22c, the surface can be deliberately textured by a conventional patterning technique, for example, by nanoimprint. Different from self-ordered two-step anodic anodization which only generates circular-shaped pores with close-packed honeycomb (i.e., trigonal) arrangement, artificially nanoengineering Al foil surface by hard imprint lithography can guide anodic anodization to form pores with desired structural parameters. Nanopores will be formed with controlled regularity and density. The pore size and spacing can be further regulated by unequal anodization rates at different anodization voltages [110]. An example is shown in Fig. 10.23, where nanoimprint was used to pattern the aluminum surface and subsequent anodization created sparsely distributed pores instead of densely packed pores [111]. However, the anodization condition has to be tuned to match the lattice constant of surface pattern. Otherwise, no regular nanopores will be formed.

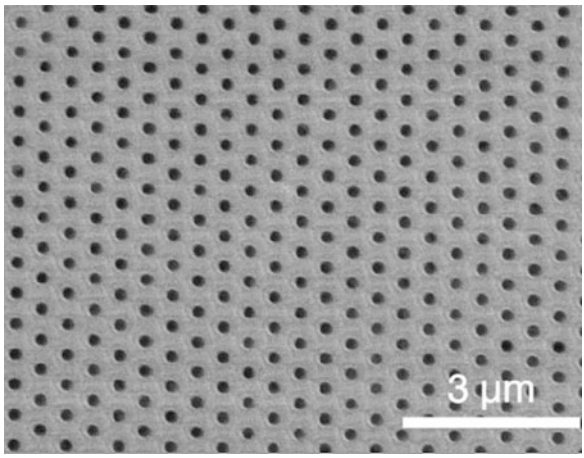


Fig. 10.23 Anodized pores of designed spacing with the assistance of nanoimprint patterning

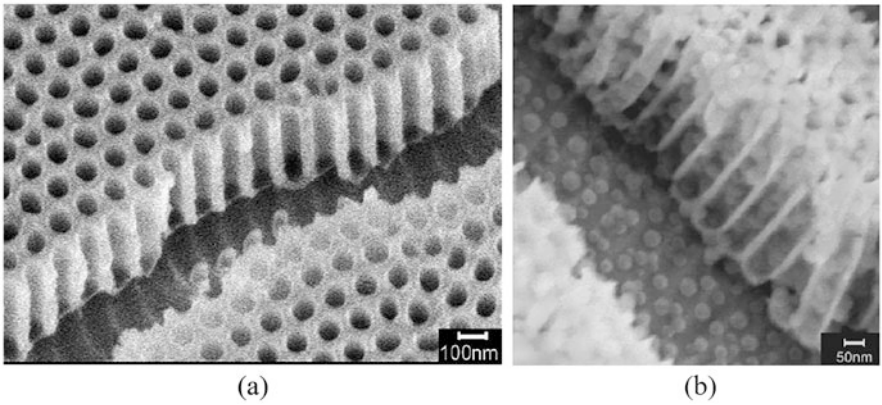


Fig. 10.24 (**a**) Cross section of ordered nanopores in alumina; (**b**) packed nanoparticles in the nanopores by electrophoretic force. (Reprinted from [117] with permission)

The most useful AAO pore for nanofabrication is to grow them in thin Al films deposited on substrates of choice, instead of bulk aluminum foils [112]. Pore sizes from 10 nm to 200 nm with aspect ratio as high as 400:1 have been made by depositing aluminum film on top of silicon and subsequent anodizing [113]. Once the silicon is used as substrate, nanostructures of other materials can be produced using the AAO nanopores as a template. For example, nanowires or nanorods can be produced by filling up the nanopores with metals [114–116]. The porous alumina templates have also been used to assemble nanoparticles, whereby nanoparticles are packed into the nanopores by electrophoretic force. Figure 10.24a shows regular nanopores created by two-step anodization of aluminum film, and Fig. 10.24b shows nanoparticles packed into the nanopores by electrophoretic force [117]. Combining with other top-down nanofabrication techniques, more exotic nanostructures have

been created [118]. One must be aware that porous alumina is an insulator, as illustrated in Fig. 10.22d. Each nanopore is surrounded by Al_2O_3. If the porous alumina is made on a conducting substrate surface, any metals deposited into the nanopores will not be in contact with the substrate surface, unless the barrier layer is removed prior to the metal deposition. The barrier layer can be dissolved in 5% phosphoric acid, albeit the acid dissolution also reduces the thickness of porous alumina layer and widens the pore size.

References

1. Whitesides, G.M., Mathias, J.P., Seto, C.T.: Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. Science. **254**, 1312 (1991)
2. Whitesides, G.M.: Self assembly and nanotechnology. In: Fourth Foresight Conference on Molecular Nanotechnology (1995)
3. Whitesides, G.M., Grzybowski, B.: Self-assembly at all scales. Science. **295**, 2418 (2002)
4. Ozin, G.A., et al.: Nanofabrication by self-assembly. Mater. Today. **12**(5), 12–23 (2009)
5. Libbrecht, K.: The enigmatic snowflakes. Phys. World. **21**(1), 19 (2008)
6. Amadi, E.V., et al.: Nanoscale self-assembly: concepts, applications and challenges. Nanotechnology. **33**, 132001 (2022)
7. Andres, R., et al.: The design, fabrication and electronic properties of self-assembled molecular nanostructures. In: Nalwa, H.S. (ed.) The Handbook of Nanostructured Materials and Nanotechnology. Academic, San Diego (1998)
8. Oliveira, O.N., et al.: The past and the future of Langmuir and Langmuir−Blodgett films. Chem. Rev. **122**(6), 6459–6513 (2022)
9. Ulman, A.: Formation and structure of self-assembled monolayers. Chem. Rev. **96**, 1533–1554 (1996)
10. Xia, Y., Whitesides, G.M.: Soft lithography. Angew. Chem. Int. Ed. **37**, 550–575 (1998)
11. Motesharei, K., Myles, D.C.: Molecular recognition on functionalized self-assembled monolayers of Alkanethiols on gold. J. Am. Chem. Soc. **120**, 7328–7336 (1998)
12. Jeoung, E., Carroll, J.B., Rotello, V.M.: Surface modification via 'lock and key' specific self-assembly of polyhedral oligomeric silsequioxane (POSS) derivatives to modified gold surfaces. Chem. Commun., 1510–1511 (2002)
13. Wenz, G., Liepold, P.: Self-assembly of biotin and thio-functionalized carboxymethyl celluloses on gold and molecular recognition of streptavidin detected by surface plasmon resonance. Cellulose. **14**(2), 89–98 (2007)
14. Lehn, J.-M.: Supramolecular Chemistry: Concepts and Perspectives. VCH (1995)
15. Aakeröy, C.B., Seddon, K.R.: The hydrogen bond and crystal engineering. Chem. Soc. Rev. **22**(6), 397 (1993)
16. Stupp, S.I., et al.: Supramolecular Materials: Self-Organized Nanostructures, vol. 276, pp. 384–389. Science (1997)
17. Cheng, H., et al.: Recent advances in supramolecular self-assembly derived materials for high-performance supercapacitors. Nanoscale Adv. **5**, 2394–2412 (2023)
18. Barth, J.V., et al.: Engineering atomic and molecular nanostructures at surfaces. Nature. **437**, 671 (2005)
19. Dmitriev, A., et al.: Modular assembly of two-dimensional metal-organic coordination networks at a metal surface. Angew. Chem. Int. Ed. **41**, 2670–2673 (2003)
20. Park, J., et al.: Coulomb blockade and the Kondo effect in single-atom transistors. Nature. **417**, 722 (2002)

21. Brinker, C.J.: Evaporation-induced self-assembly: functional nanostructures made easy. MRS Bull., 631–640 (2004)
22. Machado, A.E.H., et al.: Applications of mesoporous ordered semiconductor materials - case study of TiO_2. In: Bello, S.R. (ed.) Solar Radiation Applications. IntechOpen (2015)
23. Mahoney, L., Koodali, R.T.: Versatility of evaporation-induced self-assembly (EISA) method for preparation of mesoporous TiO_2 for energy and environmental applications. Materials. **7**, 2697–2746 (2014)
24. Kimura, T.: Evaporation-induced self-assembly process controlled for obtaining highly ordered mesoporous materials with demanded morphologies. Chem. Rec. **16**, 445–457 (2016)
25. Maurin, G., et al.: The new age of MOFs and of their porous-related solids. Chem. Soc. Rev. **46**, 3104–3107 (2017)
26. Kirchon, A., et al.: From fundamentals to applications: a toolbox for robust and multifunctional MOF materials. Chem. Soc. Rev. **47**, 8611–8638 (2018)
27. Escudero, A., et al.: Molecular bottom-up approaches for the synthesis of inorganic and hybrid nanostructures. Inorganics. **9**, 58 (2021)
28. Pawel, P.: Colloidal crystals. Contemp. Phys. **24**, 25 (1983)
29. Cademartiri, L., et al.: Using shape for self-assembly. Phil. Trans. R. Soc. A. **370**, 2824–2847 (2012)
30. Rocha, B.C., et al.: Role of entropy in colloidal self-assembly. Entropy. **22**, 877 (2020)
31. Yunker, P.J., et al.: Suppression of the coffee-ring effect by shape-dependent capillary interactions. Nature. **476**, 308 (2011)
32. Denkov, N.D., et al.: Mechanism of formation of two-dimensional crystals from latex particles on substrates. Langmuir. **8**, 3183–3190 (1992)
33. Coll, A., Bermejo, S., Castañer, L.: Self-assembly of ordered silica nanostructures by electrospray. Microelectron. Eng. **121**, 68 (2014)
34. Ganai, N., Saha, A., Sengupta, S.: Colloidal particles in a drying suspension: a phase field crystal approach. The European Physical Journal E. **36**, 90 (2013)
35. Winkens, M., Korevaar, P.A.: Self-organization emerging from Marangoni and Elastocapillary effects directed by Amphiphile filament connections. Langmuir. **35**, 10799–10809 (2022)
36. Takeoka, Y.: Stimuli-responsive opals: colloidal crystals and colloidal amorphous arrays for use in functional structurally colored materials. J. Mater. Chem. C. **1**, 6059–6074 (2013)
37. Velikov, K.P., Alexander Moroz, and A.van Blaaderen, Photonic crystals of core-shell colloidal particles. Appl. Phys. Lett., 2002. 80(1): p. 49
38. Vlasov, Y.A., et al.: On-chip natural assembly of silicon photonic bandgap crystals. Nature. **414**, 289 (2001)
39. Sun, S., et al.: Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal Superlattices. Science. **287**, 1989–1992 (2000)
40. Anders, S., et al.: Lithography and self-assembly for nanometer scale magnetism. Microelectron. Eng. **61-62**, 569–575 (2002)
41. Hamdana, G., et al.: Towards fabrication of 3D isotopically modulated vertical silicon nanowires in selective areas by nanosphere lithography. Microelectron. Eng. **179**, 74–82 (2017)
42. Liddle, J.A., Cui, Y., Alivisatos, P.: Lithographically directed self-assembly of nanostructures. J. Vac. Sci. Technol. **B22**(6), 3409 (2004)
43. Chai, Z., et al.: High-rate printing of micro/nanoscale patterns using interfacial convective assembly. Adv. Mater. **32**, 2000747 (2020)
44. Juillerat, F., et al.: Fabrication of large-area ordered arrays of nanoparticles on patterned substrates. Nanotechnology. **16**, 1311–1316 (2005)
45. Ocola, L.E., Lin, X.M.: Templated self-assembly of 5nm gold nanoparticles. Proc. Nanotech Conf. **1**, 578 (2008)
46. Yin, Y., Li, Z.-Y., Xia, Y.: Template-directed growth of (100)-oriented colloidal crystals. Langmuir. **19**, 622–631 (2003)

47. Wang, D., et al.: Template-directed colloidal self-assembly—the route to 'top-down' nanochemical engineering. J. Mater. Chem. **14**, 459–468 (2004)
48. Premnath, V.A., et al.: Fabrication of three-dimensional opal nanolattices using template-directed colloidal assembly. J. Vac. Sci. Technol. B. **40**, 062803 (2022)
49. Tokuno, T., et al.: Transparent electrodes fabricated via the self-assembly of SilverNanowires using a bubble template. Langmuir. **28**(25), 9298–9302 (2012)
50. Guo, M., et al.: Advanced unconventional techniques for sub-100 nm nanopatterning. InfoMat. **4**, e12323 (2022)
51. Law, K.-Y.: Definitions for hydrophilicity, hydrophobicity, and Superhydrophobicity: getting the basics right. J. Phys. Chem. Lett. **5**(4), 686–688 (2014)
52. Jokinen, V.: Hydrophilic and hydrophobic patterning. In: Li, D. (ed.) Encyclopedia of Microfluidics and Nanofluidics. Springer (2020)
53. Fan, F., Stebe, K.J.: Assembly of colloidal particles by evaporation on surfaces with patterned hydrophobicity. Langmuir. **20**, 3062–3067 (2004)
54. Minari, T., et al.: Room-temperature printing of organic thin-film transistors with π-junction gold nanoparticles. Adv. Funct. Mater. **24**, 4886–4892 (2014)
55. Krieger, A., Zika, A., Gröhn, F.: Functional Nano-objects by electrostatic self-assembly: structure, switching, and Photocatalysis. Front. Chem. **9**, 779360 (2022)
56. Tien, J., Terfort, A., Whitesides, G.M.: Microfabrication through electrostatic self-assembly. Langmuir. **13**, 5349–5355 (1997)
57. McCarty, L.S., Winkleman, A., Whitesides, G.M.: Electrostatic self-assembly of polystyrene microspheres by using chemically directed contact electrification**. Angew. Chem. Int. Ed. **46**, 206–209 (2007)
58. Trau, M., Saville, D.A., Ksay, I.A.: Field-induced layering of colloidal crystals. Science. **272**, 706–709 (1996)
59. Trau, M., Saville, D.A., Ksay, I.A.: Assembly of colloidal crystals at electrode interfaces. Langmuir. **13**, 6375–6381 (1997)
60. Solomentsev, Y., BoÈhmer, M., Anderson, J.L.: Clustering and pattern formation during electrophoretic deposition: a hydrodynamic model. Langmuir. **13**, 6058–6068 (1997)
61. Hayward, R.C., Saville, D.A., Aksay, I.A.: Electrophoretic assembly of colloidal crystals with optically tunablemicropatterns. Nature. **404**, 56 (2000)
62. Lee, J., Isobe, T., Senna, M.: Preparation of ultrafine Fe_3O_4 particles by precipitation in the presence of PVA at high pH. J Colloid and Interface Science. **177**, 490–494 (1996)
63. Liu, N., et al.: Progress on rapidly and self-assembly magnetically responsive photonic crystals with high Tunability and stability. Front. Mater. **9**, 843097 (2022)
64. Yellen, B.B., Friedman, G.: Programmable assembly of colloidal particles using magnetic micro-well templates. Langmuir. **20**, 2553–2559 (2004)
65. Rothemund, P.W.K.: Folding DNA to create nanoscale shapes and patterns. Nature. **440**, 297–302 (2006)
66. Majikes, J.M., Liddle, J.A.: DNA origami design: a how-to tutorial. J. Res. Natl. Inst. Stand. Technol. **126**, 126001 (2021)
67. Sharma, A., et al.: DNA nanostructures: chemistry, self-assembly, and applications. In: Barhoum, A., Makhlouf, A. (eds.) Emerging Applications of Nanoparticles and Architecture Nanostructures. Elsevier (2018)
68. Vittala, S.K., Han, D.: DNA-guided assemblies toward Nanoelectronic applications. ACS Applied Bio Materials. **3**(5), 2702–2722 (2020)
69. Michelson, A., et al.: Three-dimensional nanoscale metal, metal oxide, and semiconductor frameworks through DNA-programmable assembly and templating. Sci. Adv. **10**(2), eadl0604 (2024)
70. Cai, Z., et al.: From colloidal particles to photonic crystals: advances in self-assembly and their emerging applications. Chem. Soc. Rev. **50**, 5898–5951 (2021)
71. Colson, P., et al.: Nanosphere lithography: a powerful method for the controlled manufacturing of nanomaterials. J. Nanomaterials. **2013**, 948510 (2013)

72. Laurvick, T.V., et al.: Surface feature engineering through nanosphere lithography. J. Micro/Nanolith. MEMS MOEMS. **15**(3), 031602 (2016)
73. Ai, B., Zhao, Y.: Glancing angle deposition meets colloidal lithography: a new evolution in the design of nanostructures. Nano. **8**(1), 1–26 (2019)
74. Yabagi, J.A., et al.: Nanofabrication process by reactive ion etching of polystyrene Nanosphere on silicon surface. J. Sci. Technol. **9**(3), 145–153 (2017)
75. Chen, J., et al.: Controllable fabrication of 2D colloidal-crystal films with polystyrene nanospheres of various diameters by spin-coating. Appl. Surf. Sci. **270**, 6–15 (2013)
76. Wu, W., et al.: A deep sub-wavelength process for the formation of highly uniform arrays of nanoholes and nanopillars. Nanotechnology. **18**, 485302 (2007)
77. Zhang, Z., et al.: Recent advancement on micro−/nano-spherical lens photolithography based on monolayer colloidal crystals. Adv. Colloid Interf. Sci. **228**, 105–122 (2016)
78. Ruzette, A.-V., Leibler, L.: Block copolymers in tomorrow's plastics. Nat. Mater. **4**, 19 (2005)
79. Leibler, L.: Theory of microphase separation in block copolymers. Macromolecules. **13**(6), 1602 (1980)
80. Park, C., Yoon, J., Thomas, E.L.: Enabling nanotechnology with self assembled block copolymer patterns. Polymer. **44**, 6725–6760 (2003)
81. Chen, Y., Xion, S.: Directed self-assembly of block copolymers for sub-10nm fabrication. Int. J. Extrem. Manuf. **2**, 032006 (2020)
82. Mansky, P., Haikin, P., et al.: Monolayer films of diblock copolymer microdomains for nanolithographic applications. J. Mater. Sci. **30**, 1987–1992 (1995)
83. Xu, H., et al.: Flow-enhanced epitaxial ordering of brush-like macromolecules on graphite. Langmuir. **22**(3), 1254–1259 (2006)
84. Morkved, T.L., et al.: Local control of microdomain orientation in diblock copolymer thin films with electric fields. Science. **273**(5277), 931–933 (1996)
85. Hahm, J., Sibener, S.J.: Cylinder alignment in annular structures of microphase-separated polystyrene-b-poly(methyl methacrylate). Langmuir. **16**(11), 4766–4769 (2000)
86. Edwards, E.W., et al.: Long-range order and orientation of cylinder-forming block copolymers on chemically nanopatterned striped surfaces. Macromolecules. **39**(10), 3598–3607 (2006)
87. Bencher, C., et al.: Self-assembly patterning for sub-15nm half-pitch: a transition from lab to fab. Proc. of SPIE. **7970**, 79700F-1 (2011)
88. Li, W., Muller, M.: Defects in the self-assembly of block copolymers and their relevance for directed self-assembly. Ann. Rev. Chem. Biomol. Eng. **6**, 187–216 (2015)
89. Liu, C.-C., Nealey, P.F.: Integration of block copolymer directed assembly with 193 immersion lithography. J. Vac. Sci. Technol. B. **28**(6), C6B30 (2010)
90. Pinto-Gómez, C., et al.: Directed self-assembly of block copolymers for the fabrication of functional devices. Polymers. **12**, 2432 (2020)
91. Chan, B.T., et al.: 28 nm pitch of line/space pattern transfer into silicon substrates with chemo-epitaxy directed self-assembly (DSA) process flow. Microelectron. Eng. **123**, 180–186 (2014)
92. Doise, J., et al.: Via patterning in the 7-nm node using immersion lithography and graphoepitaxy directed self-assembly. J. Micro/Nanolith. MEMS MOEMS. **16**(2), 023506 (2017)
93. Hinsberg, W., et al.: Self-assembling materials for lithographic patterning: overview, status and moving forward. Proc. SPIE. **7637**, 76370G-1 (2010)
94. Liu, C.-C., et al.: Towards an all-track 300 mm process for directed self-assembly. J. Vac. Sci. Technol. B. **29**(6), 06F203 (2011)
95. Welander, A.M., Kang, H., Stuen, K.O.: Rapid directed assembly of block copolymer films at elevated temperatures. Macromolecules. **41**(8), 2759 (2008)
96. Sinturel, C., et al.: Solvent vapor annealing of block polymer thin films. Macromolecules. **46**(14), 5399 (2013)
97. LaPedus, M., Sperling, E.: DSA re-enters Litho picture. In: *Semicondyctor Egngineering News*. MARCH 15TH (2018)

98. Imamura, T., et al.: Highly selective removal of poly(methyl methacrylate) from polystyrene-block-poly(methyl methacrylate) by CO/H_2 plasma etching. J. Vac. Sci. Technol. B. **33**, 061601 (2015)
99. Delalande, M., et al.: Development of plasma etching processes to pattern sub-15 nm features with PS-b-PMMA block copolymer masks: application to advanced CMOS technology. J. Vac. Sci. Technol. B. **32**(5), 051806-1 (2014)
100. Delony, J.B., et al.: Block copolymer directed self-assembly defect modes induced by localized errors in chemoepitaxial guiding underlayers: a molecular simulation study. J. Vac. Sci. Technol. B. **38**, 032604 (2020)
101. International Roadmap for Devices and Systems (IRDS™) 2022 Edition, https://irds.ieee.org/
102. Donahue, C.J., Exline, J.A.: Anodizing and coloring aluminum alloys. J. Chem. Educ. **91**(5), 711–715 (2014)
103. Masuda, H., Fukuda, K.: Ordered metal nanohole arrays made by a two-step replication of honeycomb structures of anodic alumina. Science. **268**, 1466 (1995)
104. Burgos, N., Paulis, M., Montes, M.: Preparation of Al2O3/Al monoliths by anodisation of aluminium as structured catalytic supports. J. Mater. Chem. **13**, 1458–1467 (2003)
105. Li, Y., et al.: Fabrication of highly ordered nanoporous alumina films by stable high-field anodization. Nanotechnology. **17**, 5101–5105 (2006)
106. Masuda, H., Satoh, M.: Fabrication of gold nanodot array using anodic porous alumina as an evaporation mask. Jpn. J. Appl. Phys. **35**, L126 (1996)
107. Lee, W.: The anodization of aluminum for nanotechnology applications. J Miner Met Mater Soc. **62**(6), 57–63 (2010)
108. Myung, N.V., et al.: Alumina nanotemplate fabrication on silicon substrate. Nanotechnology. **15**, 833–838 (2004)
109. Rheima, A.M., et al.: Aluminum oxide nano porous: synthesis, properties, and applications. Case Stud. Chem. Environ. Eng. **8**, 100428 (2023)
110. Xu, R., et al.: Well-defined nanostructuring with designable anodic aluminum oxide template. Nat. Commun. **13**, 2435 (2022)
111. Choi, J., et al.: Perfect two-dimensional porous alumina photonic crystals with duplex oxide layers. J. Appl. Phys. **94**(8), 4757 (2003)
112. Lee, W., Park, S.-J.: Porous anodic aluminum oxide: Anodization and templated synthesis of functional nanostructures. Chem. Rev. **114**, 7487–7556 (2014)
113. Choi, S., Daugherty, M., Walker, D.: Growth and engineering of high aspect-ratio AAO templates integrated on silicon substrates. In: Nanotech Conference, Santa Clara (2007)
114. Huber, C.A., et al.: Nanowire array composites. Science. **263**, 800 (1994)
115. Losic, D., et al.: Fabrication of gold nanorod arrays by templating from porous alumina. Nanotechnology. **16**, 2275 (2005)
116. Zaraska, L., Sulka, G.D., Jaskuła, M.: Porous anodic alumina membranes formed by anodization of AA1050 alloy as templates for fabrication of metallic nanowire arrays. Surf. Coat. Technol. **205**(7), 2432 (2010)
117. Gultepe, E., et al.: High-throughput assembly of nanoelements in nanoporous alumina templates. Appl. Phys. Lett. **90**, 163119 (2007)
118. Hu, Y., et al.: A hierarchical anodic aluminum oxide template. Nano Lett. **21**, 250–257 (2021)

Chapter 11
Applications of Nanofabrication



11.1 Introduction

The previous chapters have introduced various nanofabrication technologies. In practical applications, any fabrication technology will not be used alone. A structure or a device may not be made by one fabrication technology but a combination of several technologies. On the other hand, the same goal can be reached by different ways, and there are many factors which determine which specific technology to be used to achieve the final goal. For example, one particular technology may be superior to others in improving or advancing a specific aspect of performance in a device, or the technology is simpler than others and can lower the fabrication cost. There are also situations when fabrications have to be done using existing equipment when no other options are available or affordable.

Any new application of a nanoscale structure or device is likely to be the result of not only a novel design but also an adoption of new fabrication technology or a clever combination of different existing fabrication technologies. The developments of application and fabrication have always been interdependent to each other and advancing each other. The constant demand of new applications is the market pull to advance fabrication technologies to make something better and cheaper. The continuous progress in research of new ways of making structures or devices is the technology push to inspire designers to come up with new applications. For example, the progress of integrated circuit (IC) has been the result of continuous advance of optical lithography and associated pattern transfer technologies in the last 30 years. The applications of nanotechnology in a variety of fields have been the driving force to the development of many new nanofabrication techniques in the last 20 years.

This chapter intends to introduce the applications most relevant to the advances in nanofabrication technologies, including integrated circuits, photonics, and nanotechnology; describes how different fabrication techniques or processes are used in each field; and cross-references them to the previous chapters of this book. These

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6_11

applications have best manifested how modern nanofabrication technologies play an important role and how fabrication and application rely on each other as well as advancing each other. It should be noted that nanofabrication cannot stand alone. It evolves from microfabrication and is often being used in combination with microfabrication. The boundary between microfabrication and nanofabrication is blurry, and this chapter will focus on those applications which require mostly the fabrication of sub-100 nm features or structures.

11.2 Integrated Circuits

Gordon Moore, one of the founders of Intel, famously predicted in 1965 that the number of transistors on a silicon chips would double every 18 months, which late in 1975 was modified to every 2 years [1]. The "Moore's law" has been more or less right in nearly 60 years, although the rate of doubling in transistor density was slowed down from 24 months to 33 months as shown in Fig. 11.1 [2]. In recent years, Moore's law no longer implies simple shrinking of transistor dimensions. In fact, transistor itself has evolved from traditional 2D planar structure in Fig. 11.2a to 2.5D fin-like structure, such as the FinFET shown in Fig. 11.2b, because the gate electrode simply cannot effectively control the flow of electrons on one side of the semiconductor channel as its length shrinks narrower and narrower. A FinFET has

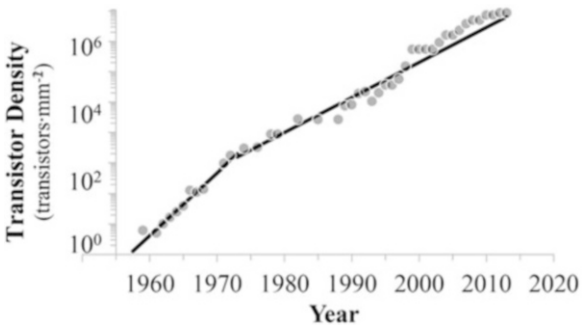


Fig. 11.1 The transistor density evolution of Intel processors by year [2]

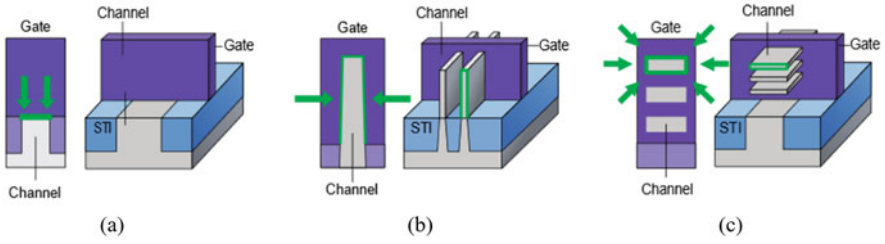


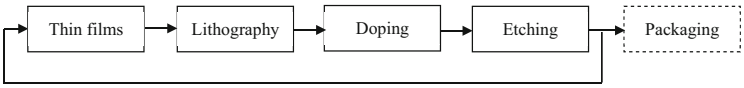
Fig. 11.2 (**a**) Planer FET (field-effect transistor), (**b**) FinFET, (**c**) GAA FET [4]

its gate surrounding three sides of the channel, resulting in greatly enhanced controllability in that even the channel becomes narrower. From 22 nm technology node onward, all transistors in an IC chip have been FinFETs. It is the FinFET which saved the Moore's law [3]. Further shrinking to below 7 nm technology node, even FinFET can no longer function efficiently. The transistor goes to 3D with the gate around four sides of a channel, as shown in Fig. 11.2c which is the gate-all-around transistor (GAA FET).

The aforementioned "technology node" which corresponds to the minimum feature size of transistors originally came from the "National Technology Roadmap for Semiconductors" (NTRS). The first edition of NTRS was published in 1992, which was based on the semiconductor industrial consensus that the transistor would shrink in size so that more transistors could be integrated in accordance to the Moore's law. In 1999, NTRS became ITRS (International Technology Roadmap for Semiconductors) with the inclusion of worldwide semiconductor players. In the 2001 edition of ITRS, the technology node was redefined by minimum half-pitch (hp) size instead of minimum feature size (see definition of half-pitch in Sect. 2.7 of Chap. 2). In 2016, ITRS evolved to become IRDS (International Roadmap for Devices and Systems) to reflect the wide spectrum of electronics not just ICs [5].

One should be aware that the technology node which explicitly indicates the minimum half-pitch size of planar transistors is no longer valid after the 22 nm node, because of the evolution of transistor architecture to FinFET and GAA FET (Fig. 11.2). For example, the 5 nm node of Samsung's IC actually has a minimum metal pitch of 36 nm, corresponding to the 18 nm (half-pitch) technology node definition according to NTRS/ITRS and IRDS nomenclature. This means that the semiconductor industry is in the 18 nm generation in 2020 if using the correct definition of technology node. However, IC manufacturers still insist in calling this technology "the 5 nm node" even though these node numbers do not correspond to any meaningful and measurable quantity related to transistor density on the wafer. For this reason, it is apparent that Moore's law will still be valid for the next 10 years because the real sizes of IC features are far bigger than the node definition [6].

To be able to cram more and more transistors into a single silicon chip is entirely the result of continuous advances in fabrication technologies. Making IC chips is an enormous piece of system engineering [7]. It involves more than a thousand process steps, including materials, design, processing, packaging, and test. The essence of IC manufacturing is the planar technology, i.e., an IC structure, no matter how complicated, is built layer upon layer from a substrate material up. The processes can be grouped into four basic cycling stages before final packaging:



There are two major types of thin film: conductive thin film and insulating thin films. Conductive thin films form the circuit interconnection system, and insulating thin films separate different circuit functions. The methods for thin film deposition

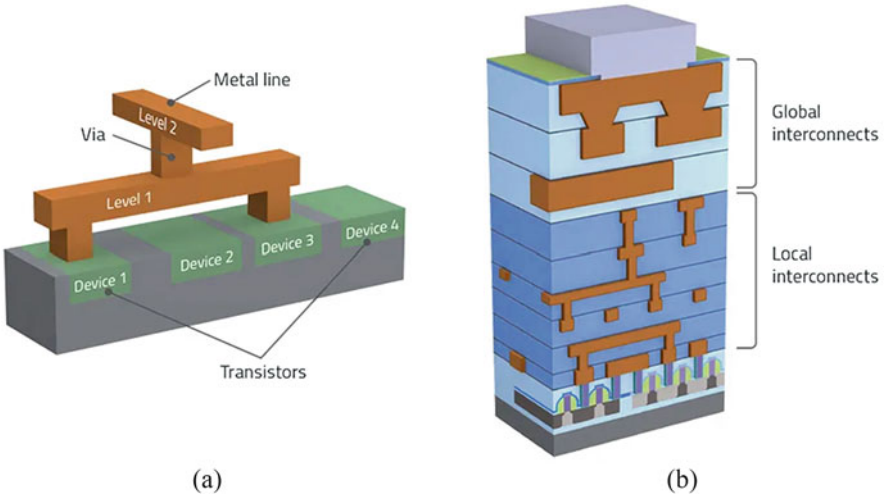


Fig. 11.3 Interconnects of metal lines connecting transistors at local and global levels. (Courtesy of Larry Zhao, Semiconductor Engineering [8])

include thermal evaporation, ion sputtering, chemical vapor deposition, and molecule beam epitaxy. Doping forms the semiconducting channel of a transistor, which is the most basic element of ICs. The techniques for doping include ion implantation and diffusion. Doping is only needed for making transistors. Once the transistors are made on the first level of processes, called front end of line (FEOL), the rest is to make interconnects, called back end of line (BEOL). Interconnects are conductive lines and holes piled up in many layers above the transistors, as illustrated in Fig. 11.3 [8]. Between layers the metal lines are connected vertically via holes (Fig. 11.3a), and the fully functional circuits are realized by both local and global interconnects (Fig. 11.3b).

For all the circuit features, the key technology which enables the making of smaller and smaller transistors is lithography. Lithography defines the photoresist patterns of IC layouts. Etching technology transfers the photoresist patterns into various thin films deposited on a silicon substrate. Lithography and etching/deposition thread through the whole IC manufacturing process. A complicated IC chip normally needs to go through 20–30 of lithography and etching/deposition steps, which is the core of IC fabrications.

As the shrinking of feature size is solely dependent on the advance of lithography technology, each edition of the semiconductor roadmap (NTRS/ITRS and IRDS) has a dedicated section to chart the progress of lithography for future generations of IC. For example, the latest roadmap (IRDS 2022 Edition) predicted the lithography options for IC manufacturing in the next 15 years as shown in Table 11.1 [9]. It should be noticed that there are two nomenclatures about the node in the table. The "Logic node" follows the tradition of semiconductor industry in planar transistor era, which continuously reduces in dimension from the 3 nm in 2022 to 0.5 nm in 2037.

Table 11.1 Potential lithography solutions for leading-edge logic ICs

	2022	2025	2028	2031	2034	2037
Logic node	1 nm	2.1 nm	1.5 nm	1.0 nm	0.7 nm	0.5 nm
Node	G48M24	G45M20	G42M16	G40M16T2	G38M16T4	G38M16T6
Minimum one-half pitch	12 nm	10 nm	8 nm	8 nm	8 nm	8 nm
Primary options for logic	EUV 0.33 NA multiple patterning	EUV 0.33 NA multiple patterning	EUV 0.55 NA single patterning	EUV 0.55 NA single patterning	EUV 0.55 NA single patterning	EUV 0.55 NA single patterning
		EUV 0.55 NA single patterning	EUV 0.55 NA multiple patterning	EUV 0.55 NA multiple patterning	EUV 0.55 NA multiple patterning	EUV 0.55 NA multiple patterning
				Beyond EUVL ($\lambda = 6.X$ nm)	Beyond EUVL ($\lambda = 6.X$ nm)	Beyond EUVL ($\lambda = 6.X$ nm)
Potential solutions for cost reduction, LER reduction		Optical +DSA EUV +DSA	Optical +DSA EUV +DSA	Optical +DSA EUV+DSA	Optical +DSA EUV+DSA	Optical +DSA EUV+DSA

However, such dimension does not have any physical meaning of a transistor on an IC chip as 0.5 nm is barely the size of a silicon atom, which is nonsense according to some expert [10]. While the semiconductor industry and general public still insist on using such nomenclature to mark the generation of ICs, the roadmap gives another definition: the Node. For 3 nm IC, the meaningful node is G40M24 which represents gate pitch of 40 nm and metal pitch of 24 nm. So the 3 nm generation of ICs have the actual 12 nm half-pitch minimum features to be patterned by lithography.

According to the roadmap (Table 11.1), the current 5 nm generation of IC is using high NA (0.33) EUV with single patterning. From 3 nm onward, multiple EUV patterning will have to be used, or higher NA(0.55) of single patterning. As the feature size gets smaller and smaller, multiple patterning using 0.55NA EUV will become the only option, or EUV with shorter wavelength (half of current EUV source) has to be developed. So far, all P-type and N-type transistors in a CMOS are made side by side on the same level. Future generations of IC will have to stack the transistors on top of each other. Table 11.1 tells that from 2028 at 1.5 nm generation ICs, the increase of transistor density is realized by two layers of transistors, which is marked by the node G40M16T2. Going to the future, even four layer (T4) or six layer (T6) transistors will be in a chip in order to keep up with the pace of development [10]. On the lithography side, EUV is the primary option whether high NA or multiple patterning. There is also a low-cost option utilizing direct self-assembly (DSA). With this option, features of smaller half-pitch are taken care of by DSA of block copolymers (Sect. 10.7 of Chap. 10), while features of large half-pitch are done by DUV multi-patterning or EUV lithography.

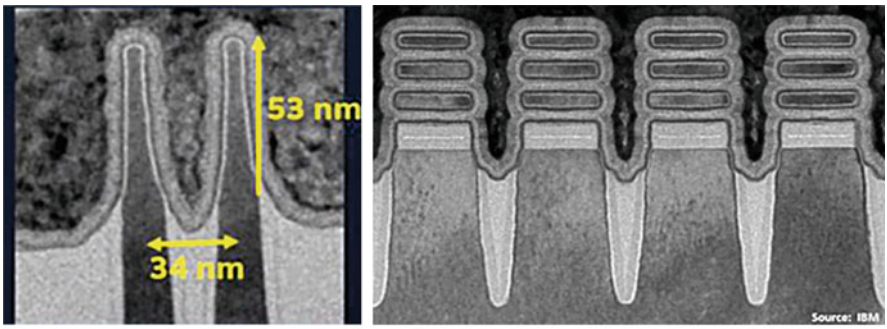


Fig. 11.4 (**a**) Intel's "10 nm" node FinFET. (Courtesy of Intel); (**b**) IBM's "2 nm" node GAA FET. (Courtesy of IBM)

One should notice that the minimum half-pitch stops at 8 nm for future generation ICs. From the nanofabrication point of view, there are already foreseeable lithography solutions for IC manufacturing in the next 15 years. The fabrication challenge is more on 3D architecture and integration. Looking at the transistor structures of Intel's "10 nm" node FinFET [11] and IBM's "2 nm" node GAA FET [12] shown in Fig. 11.4, the fabrication challenges are not only about lithography but also about high aspect ratio etching and highly conformable deposition in deep holes without voids. Fortunately, these technologies are already available, such as deep reactive ion etching (DRIE) and atomic layer deposition (ALD)/etching (ALE). Three-dimensional architectures are not only pursued at transistor level but also at circuit level. A typical example is the stacked flash memory circuit (NAND). The latest NAND technology is able to stack 232 layers of memory circuits in a single chip, which offers data input/output speed of 2.4 GB/s and storage density of 14.6 Gb/mm^2 [13].

While the technology nodes go down smaller and smaller, the semiconductor industry also started to take different approach since 2009, which was the "more than Moore" approach as illustrated in Fig. 11.5. The traditional shrinking of CMOS circuit feature size has continuously followed the Moore's law and has been considered as "more Moore" approach in Fig. 11.5. The "more than Moore" approach is to integrate diverse functionalities with the traditional ICs. There two ways to realize more than Moore, which are the system on chip (SoC) and system in package (SiP).

Conventional ICs are classified as CPU, DRAM/flash, ASIC, digital, or analog circuits. There are also microelectromechanical systems (MEMS) and various silicon-based sensors. The concept of SoC is to combine these different functional circuits or sensors into a single chip, so that the chip itself becomes an integrated system. SoC allows much smaller and compact electronic systems with multiple functions. There are many challenges to realize SoC because of the different types of circuits, particularly the MEMS type of sensors [14]. They require different fabrication processes and different design methodologies. To put them all together may

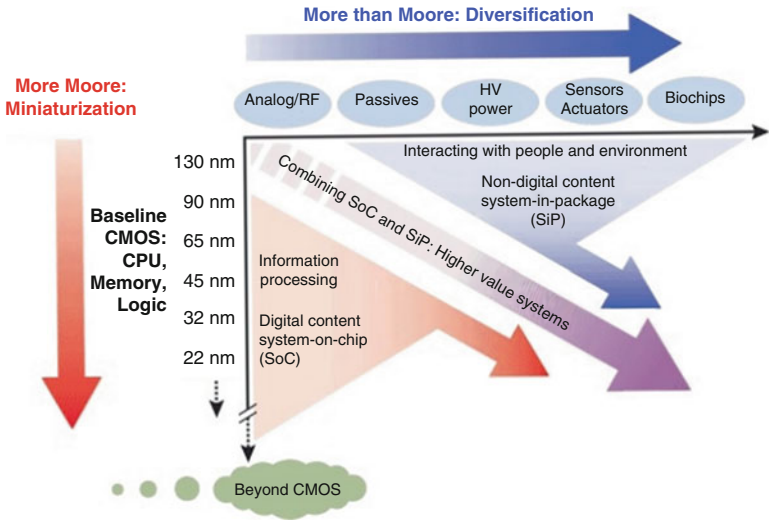
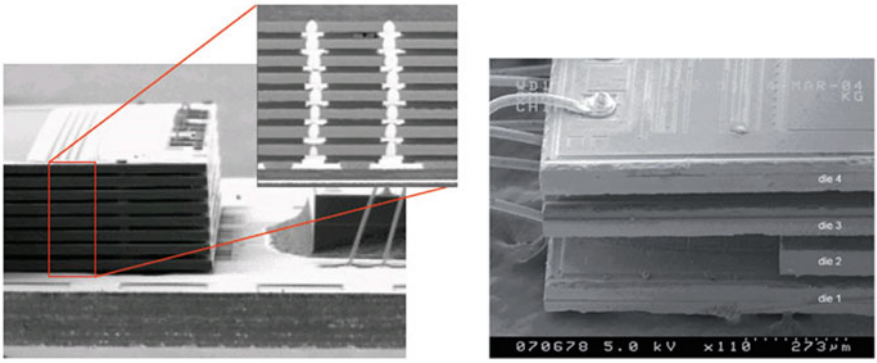


Fig. 11.5 More Moore and more than Moore



(a) (b)

Fig. 11.6 (**a**) SiP by through silicon via (TSV) [18], (**b**) SiP by wire bounding [19]

introduce interferences or reduce manufacturing yield, and the manufacturing cost is high because of incompatibility of different fabrication technologies. System in package (SiP) takes a different approach from SoC. Instead of putting multifunctions into a single die, SiP put different dies with different functions into a single packaged chip. Conventional IC package only contains a single die (chip). SiP is a single chip consisting of multiple dies stacking on top of each other, or side by side sitting on a silicon interposer, instead of different chips on a printed circuit board (PCB) [15]. The interconnection between stacked dies can be realized in two ways: one is

by through silicon via (TSV) and the other is by wire bonding or flip-chip bonding. Examples of the two approaches are shown in Fig. 11.6. To put multiple dies into a single package, each die has to be thin down from its original a few hundreds of micrometers to a few tens of micrometers. The TSV is achieved by deep silicon etching which has been introduced in Chap. 7.

The third approach is chiplets [16]. The idea behind chiplets is to break apart the system on a chip (SoC) into its composite functional blocks, or parts. Sub-elements of a complex-function chip could be made as chiplets (small chips), where these sub-elements may include separate computational processor or graphics unit, an AI accelerator, an I/O function, or a host of other chip functions. A system comprised of chiplets is kind of like an SoC on a module. Chiplet technology can help extend Moore's law by allowing the creation of more complex and powerful SoCs without having to fit all of the necessary components onto a single, monolithic chip. Chiplet-based architectures offer several benefits over traditional monolithic SoCs, including improved performance, reduced power consumption, and increased design flexibility. The chiplets can be packaged in the so-called "CoWoS" (Chip-on-Wafer-on-Substrate) scheme, which is a 2.5D wafer-level multi-chip packaging technology incorporating multiple dies, particularly processors and memories, side by side on a silicon interposer to achieve better interconnect density and performance [17].

In early 2016, an article published in *Nature* pointed out that by 2020 the transistor will only be 2–3 nm according to the Moore's law, which is just ten atoms across. At that scale, electron behavior will be governed by quantum uncertainties that will make transistors hopelessly unreliable. The article then declared that the Moore's law could not go on forever [20]. Now we know that transistors do not have to be that small. Even in 15 years from now, the minimum feature size in a transistor will still be 8 nm, as the IRDS reveals (Table 11.1). The scientific and technology communities of semiconductors used to worry about the physical limit of CMOS integrated circuits. Alternatives such as single-electron transistors or non-FET not based on charge transportation phenomena were proposed and investigated [21]. With the transistors going to 3D, the physical limit is no longer a key issue. The Moore's law also evolved to not a simple count of transistors, but comprehensive scaling of Power, Performance, Area, Cost (PPAC) by which the transistor size is only one of the parameters (area). The new IC scaling of Moore's law sets the targets as follows: for every 2–3 years, $>10\%$ more operating frequency at scaled supply voltage (Performance); $>20\%$ less energy per switching at a given performance (Power); $>30\%$ less chip area footprint (Area); and $< 30\%$ more wafer cost or 15% less die cost for scaled die (Cost) [22].

To keep up with the new Moore's law, simple reduction of transistor size is no longer enough. Figure 11.7 shows the evolution of MOSFET architecture since the FinFET [23]. From FinFET to GAA FET or nanosheet FET (NSFET), the controllability of single transistor is greatly enhanced. However, a CMOS is composed of both pFET and nFET. So far, they are placed side by side on the same level as the Forksheet FET in Fig. 11.7c. Further scaling has to reduce the channel width, hence the reduction of transistor performance. The next generation of ICs is to stack up the P-type and N-type transistors, which is the complementary FET (CFET) shown in

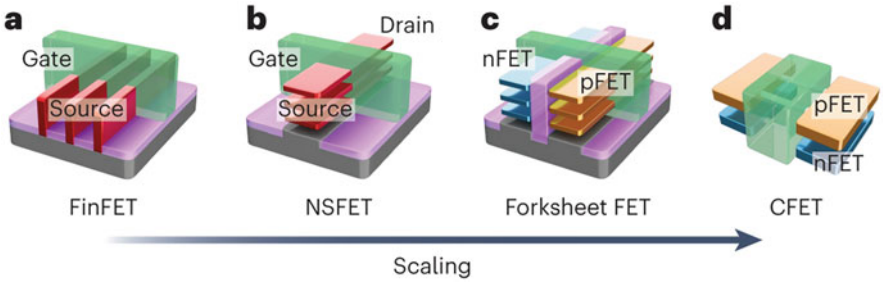


Fig. 11.7 Architecture evolution of MOSFETs. (**a**) FinFET. (**b**) NSFET. (**c**) Forksheet FET with horizontally spaced nFET and pFET structures. (**d**) CFET with vertical nFET and pFET stacking [23]

Fig. 11.7d. CFET allows maximization of the effective channel width without sacrificing the chip area, hence the enhancement of transistor's drive current and working speed. From the fabrication point of view, two possible schemes are being explored for enabling the vertical stacking: monolithic and sequential. A monolithic CFET starts with the epitaxial growth of the bottom channel and followed by the deposition of an intermediate sacrificial layer. Next comes the epitaxial growth of the top channel [24]. Although this seems the most straightforward way to build CFETs, the stacking approach results in the fabrication of very high aspect ratio vertical structures. The fabrication of second layer transistors requires high annealing temperature (900 °C) for doping diffusion, which can harm the transistors underneath. To avoid degradation of the bottom tier structures, the overall thermal budget of the top tier device as well as the layer transfer processing needs to remain low, typically below 500 °C. The sequential approach relies on bonding a blanket pre-doped single-crystalline wafer onto the bottom processed wafer prior to pursuing top tier fabrication process [25].

11.3 Photonics

Photonics is the science and technology of light [26]. It encompasses generating, guiding, manipulating, amplifying, and detecting light. Lasers, optical fibers, cameras and screens in mobile phones, lighting in cars and homes, computer screens, and TVs are just a few examples of photonics. Photonics plays an important role in driving innovation across an increasing number of fields from optical data communications to imaging, lighting and displays, manufacturing sector, life sciences, healthcare, security, and safety. It is behind many of the innovations which have transformed the way we live in recent years [27]. The global market for light-enabled systems is about $1.4 trillion today and is expected to reach nearly $2 trillion by

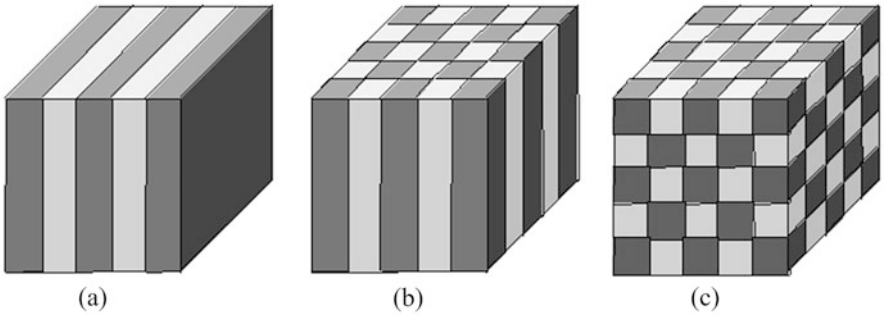


Fig. 11.8 Illustration of (**a**) 1D, (**b**) 2D, and (**c**) 3D photonic crystals

2025, in which the photonics components represent about 9% of this total share, or roughly $120 billion [28].

One of the most basic photonic devices/structures/materials is photonic crystals which can be 1D, 2D, or 3D, as illustrated in Fig. 11.8. Photonic crystals are also called photonic bandgap materials. It is well known that light is refracted or reflected at the interface of two different mediums. If one material is periodically embedded in another material as shown in Fig. 11.8, and each of them has different refractive index, light traveling inside the material will experience refraction and reflection by the periodically distributed structures, resulting in a light wave of a particular energy band completely blocked from going through the material, or a bandgap is formed for a particular wavelength of light. It is very similar to electrons scattering in a semiconductor crystal where periodical arranged atoms form different energy bands to the free traveling electrons. Therefore, photonic crystal is an electromagnetic wave transmitting material which is periodically embedded with a material of another refractive index [29].

The concept of photonic crystal was first proposed in 1987 by researchers in Bell Labs [30]. The first photonic crystal was a piece of microwave transmission material with 1-mm-diameter periodical holes fabricated by mechanical drilling. The drilled material was able to block microwave transmission in any direction in the material. The practical applications of photonic crystals in near infrared (wavelength 780–3000 nm) and visible light (wavelength 450–750 nm) emerged 10 years later. For optical wavelength, the periodical structures have to be in sub-micrometer dimension, which was only possible by the progress of fabrication technology at the time.

It must be pointed out that making photonic crystals is not just for blocking electromagnetic wave transmission but to control the transmission in desirable fashion, for example, to guide optical wave to travel along designed routes. To enable this application, the photonic crystal must contain designed defects. The so-called defects are the region where there are no periodical structures, therefore no blocking of waves. One of such examples is shown in Fig. 11.9 where apart from the periodical holes, there is a path without the holes [31]. Optical wave is blocked

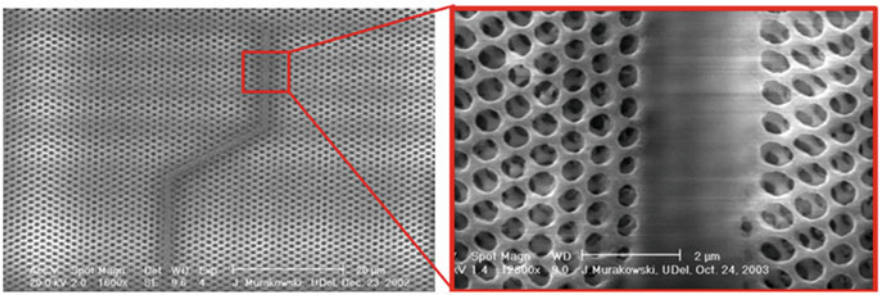


Fig. 11.9 Example of optical wave guide realized by 3D photonic crystal structure

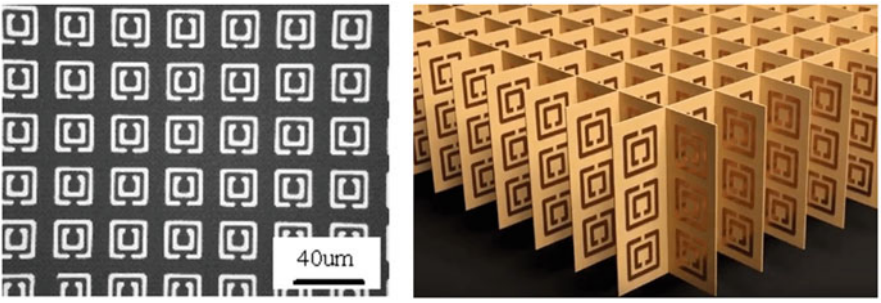


Fig. 11.10 Example of 2D and 3D metamaterials (split ring resonators)

by the photonic bandgap and only allowed to travel along the hole-free path which forms the wave guide. Introducing defects into a photonic crystal is not straightforward, and a careful design of fabrication process is required. Many fabrication technologies introduced in previous chapters can be employed to make photonic crystals, including conventional top-down lithography and pattern transfer [32, 33], as well as bottom-up self-assembly of colloid particles and nanospheres [34] or block copolymers [35, 36]. For example, the 3D photonic crystal waveguide shown in Fig. 11.9 was fabricated by a combination of UV interferometric (holographic) lithography and electron beam lithography in thick photoresist utilizing the depth control by electron energy [32].

Another class of manmade materials, similar to photonic crystals, is metamaterials which are engineered so that a property not found in naturally occurring materials could be exploited. Metamaterials consist of 2D or 3D periodical metallic or dielectric structures which are of negative index of refraction to incident electromagnetic waves. Classic refraction of electromagnetic wave through a medium material is right-handed. The refracted wave through metamaterials is however left-handed. The change of refraction direction has the possibility of making an object invisible [37], which has been the main drive to study this type of materials. The classic configuration of metamaterials is metallic ring type of

structures on an insulating surface, called split ring resonators (SRRs), as shown in Fig. 11.10 [38]. The dimension of SRR structures depends on the electromagnetic wavelength. The shorter the wavelength, the smaller the rings should be. The SRRs shown in Fig. 11.10 were for terahertz and far infrared wavelength. The fabrication of SRR is quite straightforward, basically lithographic patterning and metal liftoff. For small SRR, e-beam lithography can be used for patterning [39].

In recent years, the concept of metamaterial has evolved to metasurfaces which are thin 2D metamaterial layers that allow or inhibit the propagation of electromagnetic waves in desired directions. The main advantages of metasurfaces compared to the existing conventional technology are their low cost, low level of absorption in comparison with bulky metamaterials, and easy integration due to their thin profile [40]. Metasurfaces manipulate electromagnetic waves by scattering from the nanostructured antennas on the surface, which enable anomalous reflection or transmission, in contrast to conventional elements such as lenses. They are one of the most promising candidates for replacing bulk optical components and devices, for example, the optical lens in virtual reality (VR) headsets. The propagation characteristics of a wave that reflects from a metasurface depend strongly on the physical shapes which are of sub-optical wavelength dimension and arrangement of these antennas [41]. The key issue for widespread application of metasurfaces is the large scale and high-throughput fabrication of metasurfaces. Although optical and e-beam lithography have been employed to fabricate metasurfaces, other low-cost high-throughput technologies have also been explored such as roll-to-roll nanoimprint [42, 43] and nanosphere lithography [41].

Microoptical elements are another application domain where nanofabrication can play a key role. The microoptical elements include refractive and diffractive microlenses, light diffusion, guiding, coupling or antireflection, and anticounterfeiting surfaces. This type of applications requires surface relief structures on quartz, glass, or plastic films with dimensions equivalent to optical wavelength. The fabrication techniques involved include mostly lithography patterning and reactive ion etching. To make large-area surface relief structures on plastic film, roll-to-roll nanoimprint is commonly used. As introduced in Chap. 6, nanoimprinting lithography is particularly suitable for fabricating this type of structures which are dimensionally uniform, periodically distributed, and only requiring single layer patterning.

One particular application which has drawn wide attention is the patterned sapphire substrate (PSS) for light emitting diode (LED). Conventional LEDs have the light emission layer sandwiched between a sapphire substrate and a transparent electrode. Because of the difference of refractive index for these layers, light emitted from the emission layer may undergo total reflection between the transparent electrode and sapphire substrate and be trapped inside, as schematically shown in Fig. 11.11a, resulting in low light extraction efficiency. If the sapphire substrate surface is not flat but rugged, the total reflection could be interrupted, and light extraction efficiency may be improved, as shown in Fig. 11.11b. There are many techniques to make the sapphire substrate surface rugged, such as wet chemical etching [44, 45] and reactive ion etching [46].

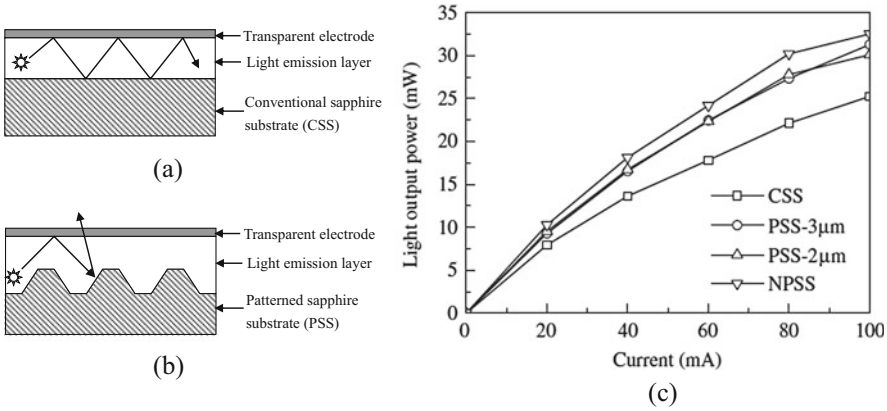


Fig. 11.11 (**a**) Conventional sapphire substrate causing light trapping. (**b**) Rugged sapphire substrate increasing light extraction. (**c**) Comparison of light output of LEDs with conventional sapphire substrate (CSS) and patterned sapphire substrate (PSS) with 3-μm-diameter bumps, 2-μm-diameter bumps, and 450-nm-diameter bumps (NPSS)

The ruggedness can be controlled by patterning the surface with regular patterns. Figure 11.11c compares the light output for conventional sapphire substrate (CSS) and patterned sapphire substrate (PSS) with different pattern structures [47]. It demonstrates that the PSS can indeed enhance the light output and the best extraction is achieved with nanoscale patterned sapphire substrate (NPSS) or nano-micro complex PSS [48]. Many patterning techniques have been employed to create the PSS, such as optical lithography [49], nanoimprint [50], and self-assembly [51]. The shapes of the patterned surface relief structures can also be varied as pillars, hemispherical and conical, etc. [52]. The improvements in light extraction efficiency are also varied from 20% to 60%. However, it is hard to draw any firm conclusion about which type or structural parameter of PSS is the best for LED applications [53]. PSS has been widely employed in manufacturing for LEDs, and making PSS has become an industry itself. Not only LEDs, organic light emitting diodes (OLEDs) and solar cells also started adopting the PSS concept to improve either light extraction in the case of OLEDs or light adsorption in the case of solar cells.

11.4 Nanotechnology

Nanotechnology is a very broad technology field. If nanotechnology can be summarized in one sentence, that is, "Nanotechnology is the understanding and control of matter at dimensions between approximately 1 and 100 nanometers (nm), where unique phenomena enable novel applications not feasible when working with bulk materials or even with single atoms or molecules" (quote from the US National Nanotehnology Initiative [54]). The first person who predicted nanotechnology was

Dr. Richard Feynman, the Nobel Laureate in Physics. In his famous speech in 1959, he predicted "there is plenty of room at the bottom," meaning that many things remained to be discovered at the most basic level of matters. By the 1980s, the development of scanning tunneling microscope (STM) and atomic force microscope (AFM) enabled humankind to see and touch the atomistic world and to manipulate atoms. The discovery of carbon nanotubes in 1991 opened the world of nanomaterials and demonstrated their amazing properties. Since 2000, marked by the publication of National Nanotechnology Initiative by the US government, nanotechnology became the focus of worldwide scientific and technology community. By the end of 2023, the US government has invested more than 40 billion US dollars in the R&D of nanotechnology [55]. With rapid advances in material science and fabrication technologies, the particularity of nanoworld becomes more and more evident. Material structures below 100 nm, either naturally formed or manmade, have displayed completely different properties from their bulk form. These properties have resulted in many applications previously impossible with bulk materials or with microscale structures. The global nanotechnology market was valued at $57 billion in 2022 and will reach $183 billion by 2028 [56]. Nanotechnology encompasses many areas of applications. While nanofabrication itself is one of the subsets of nanotechnology, there are other application fields including nanomaterials, nanoelectronics, nanophotonics, and nanobioelectronics, where nanofabrication has played a key role in their development as described in the following.

11.4.1 Nanomaterials

Nanomaterials are referred to a class of materials whose entities are of 1–100 nm in physical dimension, including 0D nanomaterials such as a variety of nanoparticles and quantum dots, 1D nanomaterials such as carbon nanotubes and a variety of nanowires, and 2D nanomaterials such as graphene and a variety of 2D semiconductor materials. They are mainly inorganic such as metals, semiconductors, dielectrics, and metal oxides. While majority of nanomaterials are made by physically breaking a bulk material or chemical synthesis, there are nanomaterials which can be made by nanofabrication including previously introduced top-down and bottom-up technologies. For examples, silicon nanowires can be made by lithographic patterning of silicon substrate and subsequent etching/oxidation processes, as schematically shown in Fig. 11.12, which is a top-down approach. These silicon nanowires have found wide applications in making field-effect transistors as shown in Fig. 11.12d [57] and sensors [57, 58].

The nanowires can also be made by bottom-up approach. Using gold particles as catalysts (Fig. 11.13a), vertical silicon nanowires can be made by vapor-liquid-solid (VLS) process (Fig. 11.13b) [59]. Similar approach can be employed to grow carbon nanotubes (CNTs) at precisely controlled locations, because CNTs only grow on catalysts. Nickel as the catalyst material can be patterned by electron beam lithography into an array of dots. Vertically aligned CNTs are grown on each of the Ni dot

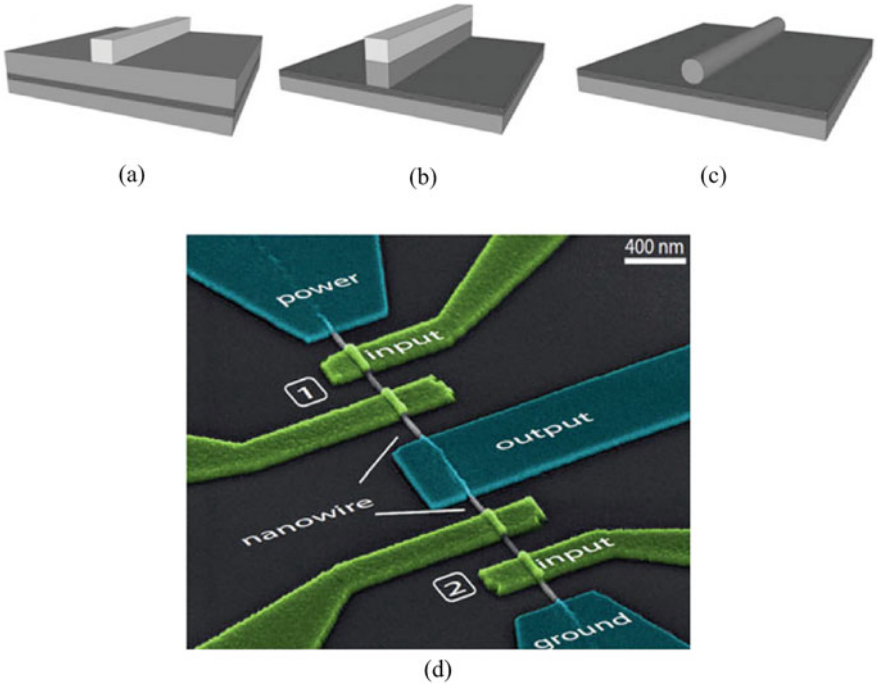


Fig. 11.12 Making silicon nanowires by top-down process. (**a**) Lithographic patterning of mask on SOI substrate. (**b**) Etching into silicon top layer. (**c**) Oxidation treatment of silicon nanowire. (**d**) Field-effect transistors made of a silicon nanowire

by PECVD process at 700 °C substrate temperature at ~150 nm/min growth rate [60]. If the Ni dot is sufficiently small (~100 nm), single CNT can be grown on each dot [61]. Figure 11.13c shows a single CNT grown inside a gated cavity, which serves as a field electron emitter [62]. Nanoparticles and nanowires can also be made in templates which in turn are made by nanofabrication, for example, the anodic aluminum oxide (AAO) templates [63]. Graphene is another important nanomaterial. Apart from large single layer graphene sheets which have found applications as transparent conductive films, graphene with particular shapes is needed to open up its energy bandgap. For example, graphene flakes with zigzag shape have been found with interesting spintronic properties [64]. Patterning of graphene can be done with different lithographic techniques, such as e-beam lithography, nanoimprinting, or contact printing, in combination with etching processes [65]. For the particular zigzag shape, oxygen plasma etching of circular pattern was found to be able to evolve into the zigzag shape [66].

Nanofabrication not only can create nanomaterials but also is a means to apply nanomaterials additively. A single entity of nanomaterial, though having superior property than its bulk form, is difficult to be integrated into a device or a system such as the example shown in Fig. 11.13c. The most successful industrial scale

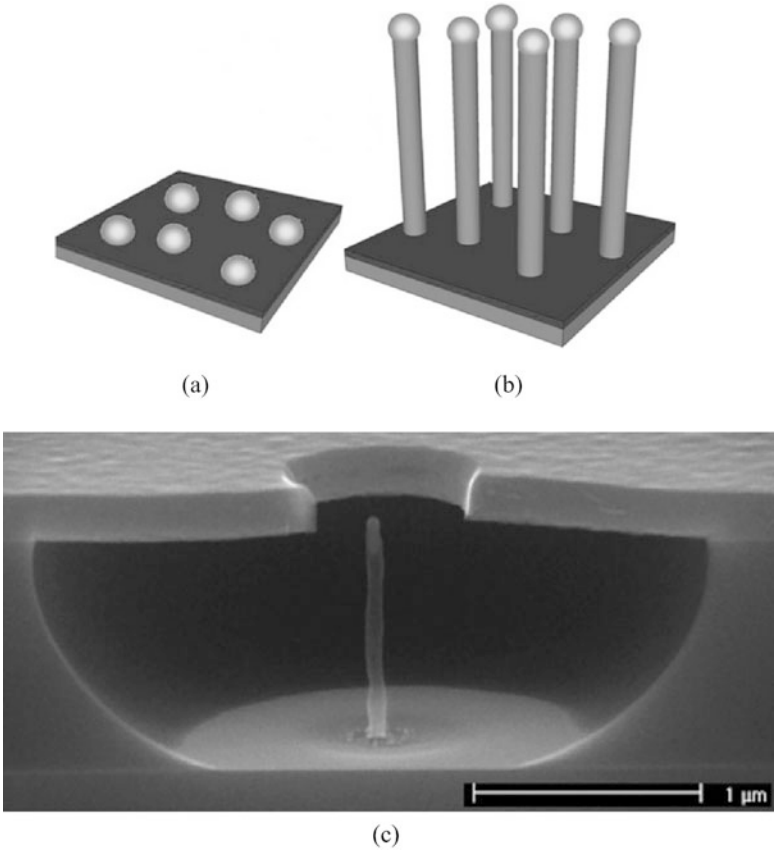


Fig. 11.13 Making silicon nanowires by bottom-up process. (**a**) Deposition of gold catalysts on silicon substrate. (**b**) VLS growth of silicon nanowires. (**c**) Single CNT grown inside a gated silicon cavity as a field electron emitter. (Reprinted from [62] with permission)

applications of nanomaterials are actually using their ink forms or powder forms and printing them [67]. Although printed pattern structures using conventional printing tools are not of nanoscale, their properties are of those of nanomaterials. One of the areas where printing nanomaterials becomes a key fabrication technology is the printed electronics [68]. Metallic nanoparticles and nanowires, graphene and carbon nanotubes, quantum dots, etc. can all be made into ink/paste and printed by inkjet, screen, and gravure printers into pattern structures or made into films by coating. Thin film transistors, light emitting diodes, and solar cells, as well as sensors and electronic circuits, can be made by printing nanomaterials. The additive nature of printing fabrication allows the electronics to be made on a variety of substrate materials, especially flexible substrates, which becomes the key technology behind the burgeoning of flexible electronics in recent years [69].

11.4.2 Nanoelectronics

In terms of physical dimension, transistors in the state-of-the-art integrated circuits are already of nanoscale as introduced in 11.2 of this chapter. Although the combination of 3D scaling and heterogeneous packaging techniques will enable the integration of a higher density of devices in a well-defined area/volume for the next two decades, as predicted by IRDS [22], the physical, topological, or technological limits will eventually be reached. The nanoelectronics in this section refers to the transistors not using traditional silicon-based material or working on field-induced charge-transport principle. In fact, researches on new paradigms of transistors and electronics have never stopped in the last 30 years.

The most investigated new paradigm is the devices based on quantum mechanical principles, notably utilizing quantum confinement effect which is promoted when the size of a material is comparable to the de Broglie wavelength of electron wave function. Size-dependent quantum confinement can also enable energy band engineering. A typical example is the single-electron transistor. Electron movement in a single-electron transistor is no longer continuous but quantized by individual electrons known as Coulomb blockade [70]. Normally such quantum confinement effect only becomes prominent at extremely low temperature (~5 K). By making ultrasmall Coulomb blockade devices with a silicon island of sub-5 nm, it can operate at room temperature (300 K). Another very promising new paradigm device is the tunnel field-effect transistor (TFET) [71]. A TFET works by modulating quantum tunneling through the source-drain barrier instead of modulating thermionic emission over the barrier as in traditional MOSFETs. In classical MOSFETs, the ability of controlling transistor current by gate voltage (Subthreshold Swing, SS) is limited to 60 mV/dec because of the Fermi-Dirac distribution which describes the probability of occupying an energy level by electrons. This hampers the reduction of the supply voltage. A considerable increase in power consumption can be expected when the device dimensions are scaled down further. For the same device dimension, TFET offers sharper turn-on devices compared to MOSFET, i.e., SS lower than 60 mV/dec, which brings the benefits of lower power consumption and higher working speed [72].

The key to the success of TFETs is to have the channel material of high charge mobility and low energy bandgap, for which silicon is not a suitable material. Since the discovery of graphene in 2004, a large family of 2D semiconductor materials with one or few atomic layer thicknesses have been developed and offered new hope for high performance TFETs. Particularly, the energy bands of these 2D materials can be engineered through nanofabrication. For example, although graphene is a promising electronic material due to its extremely high mobility, the zero bandgap limits its application in transistor devices. It has been found that nonzero bandgap can form in graphene nanoribbons which are tailored by nanofabrication [73]. Nevertheless, it should be noted that the state-of-the-art performance of the experimentally realized 2D TFETs remains less encouraging compared to their theoretical progress. This is primarily attributed to the limitation in materials and fabrication

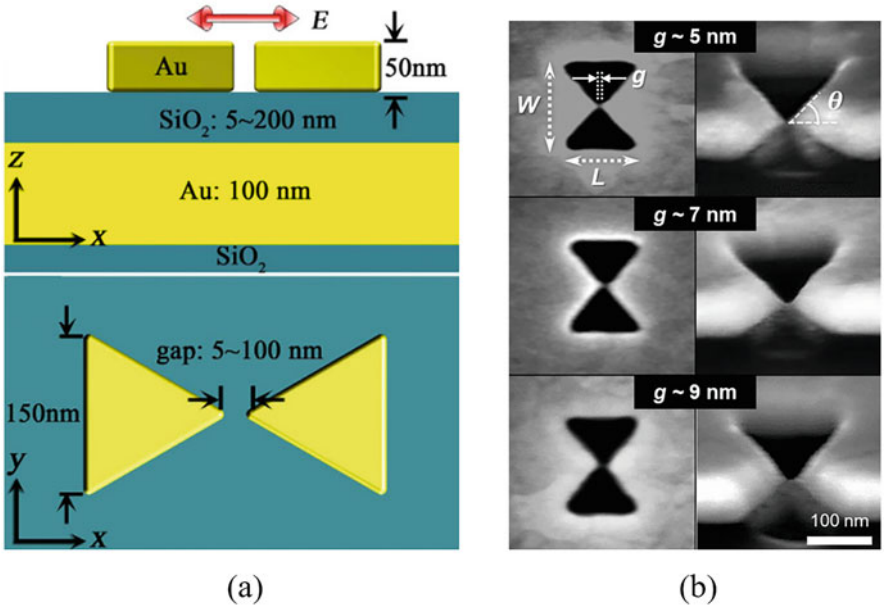


Fig. 11.14 (**a**) Schematic of bow-tie antenna which consists of two metal thin film triangles oppositely placed. (**b**) Example of sub-10 nm bow-tie structures made by FIB milling

technologies including synthesizing high-quality 2D materials with controllable layer thickness, synthesizing high-k dielectric with a high-quality dielectric/2D material interface, fabricating low-resistance contacts on 2D materials, and fabricating abrupt doping profile at tunneling junction. As the result, both average SS and on-state current in fabricated 2D TFETs so far failed to meet the desired levels needed for ensuring CMOS compatible performance [74].

11.4.3 Nanophotonics

Different from the photonics introduced in Sect. 11.3 of this chapter, which concerns light-matter interaction at the structural dimensions equivalent to optical wavelength, nanophotonics studies the flow of light at length scales far below the optical wavelength. Using metallic and dielectric nanostructures precisely sculpted into 2D and 3D nanoarchitectures, light can be scattered, refracted, confined, filtered, and processed in fascinating new ways that are impossible to achieve with natural materials and in conventional geometries [75].

One of the areas in nanophotonics, where nanofabrication is particularly relevant, is the plasmonic bow-tie antenna. The bow-tie antenna is two oppositely placed metal triangle pieces as schematically shown in Fig. 11.14a [76]. When the bow-tie

antenna is illuminated with a laser light polarized to gap direction, it excites the plasmon resonance at the apex between metal triangles. The plasmonic resonance has been introduced in 2.8 of Chap. 2. It is a phenomenon of free electrons in a thin metal sheet becoming excited by light, which generates high intensity electric fields. The resonance intensity can be from several hundred to several ten thousand times higher than that of the incident light. The narrower the bow-tie gap, the stronger the plasmon resonance. For example, the enhancement factor enabled by a 3 nm gap was one order of magnitude higher than that of a 50 nm gap for surface-enhanced infrared spectroscopy [77]. To fabricate such narrow metallic gaps, many techniques can be used, including direct e-beam lithography, focused ion beam sculpturing or nanoimprinting [73], as well as a variety of indirect methods as described in 9.5 of Chap. 9. Figure 11.14b shows an example of bow ties with gaps of 5, 7, and 9 nm made by a Ga^+ based focused ion beam milling [78]. By FIB induced grain boundary breaking, the gap can be made as small as 1–2 nm [79]. In fact, the entire field of nanophotonics is dependent on the advances in nanofabrication technologies. Nanofabrication is the "unsung hero" in enabling advances in nanophotonics [80].

11.4.4 Nanobiotechnology

Nanobiotechnology is a discipline in which tools from nanotechnology are developed and applied to study biological phenomena. One of the most important tools from nanotechnology is nanofabrication, which creates nanoscale structures in a variety of ways to help the progress of biotechnology. Although majority of biological applications in nanobiotechnology are utilizing wide spectrum of nanomaterials [81], nanoscale devices are also important helpers which crucially relies on nanofabrication. One such category is nanoscale biosensors employing nanoelectromechanical systems (NEMS). In NEMS, the key nanoscale entity is a mechanically movable element or structure which is a free standing thin beam, either fixed at both ends (bridge) as shown in Fig. 11.15a [82] or fixed at one end (cantilever) as shown in Fig. 11.15b [83]. They can be driven electrically into vibration mode; therefore, they are usually called resonators. The main application of nanoresonators is as high-sensitivity biosensors to detect minute amount of biomolecules. At high resonant frequency, a tiny amount of mass increase on the resonator can change its resonant frequency, and the change can be detected. It has been reported that mass change as small as 10^{-18} g (attogram) can be detected by a cantilever nanoresonator, which is equivalent to a single DNA molecule [84]. A bridge type nanoresonator achieved mass resolution of ~zg (10^{-21} g) [85]. In recent years, nanowires/nanotubes and two-dimensional (2D) atomic layers such as graphene/phosphorene have been utilized for making nanoresonators, which offer ultralow mass, ultrawide frequency tuning range, broad dynamic range, and ultralow power consumption [86].

Fluidic platforms are essential for biology because most biological phenomena are taking place in liquid phase. While the advanced fabrication technologies

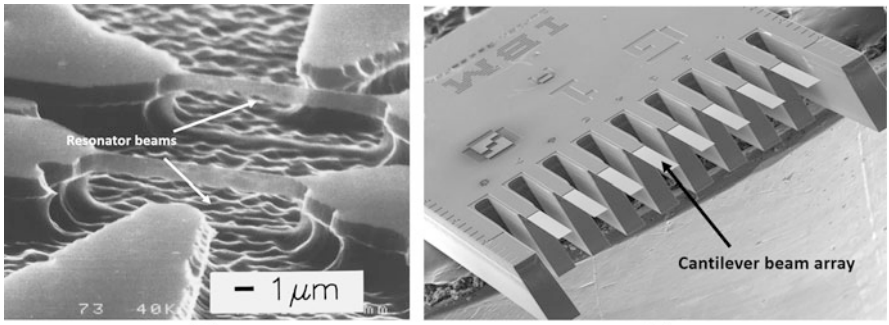


Fig. 11.15 (**a**) Bridge-type nanoresonator and (**b**) cantilever-type nanoresonator as high-sensitivity biosensors

enabled sophisticated integrated circuit (IC), the same technologies were also employed to fabricate miniaturized fluidic platforms, i.e., microfluidics, marked by the first miniaturized gas chromatograph fabricated on a silicon wafer over 40 years ago [87]. It opened the era of "lab-on-a-chip" which involves manipulating small volumes of fluids, typically in the microliter or nanoliter range, within microscale channels etched or molded on the chip [88]. Then nanofluidics emerged following the footsteps of microfluidics because of the progress in nanofabrication [89].

The three types of structures crucial to nanofluidics are nanopores, nanochannels, and nanopipets, which are all relying on state-of-the-art nanofabrication technology [90]. Many fabrication techniques introduced in previous chapters can be utilized for making nanofluidic systems [91], typically including e-beam lithography, focused ion beam milling, and etching and deposition. Where a direct nanofabrication method cannot meet the need, indirect fabrication can be employed. An example has been shown in Fig. 9.17 in Chap. 9 on how sealed nanoscale fluidic channels are made. The fabrication process is as follows: a trench is first patterned in silicon by lithography and etching processes; the trench is then etched by hydrofluoric acid (HF); subsequent atomic layer deposition (ALD) process deposits Al_2O_3 inside the tunnel walls as well as the top opening because of its conformable deposition; the deposition continues until the top opening is completely sealed off and in the same time the tunnel is narrowed down to nanometer size because of the accumulation of Al_2O_3 layer inside; eventually sub-10 nm sealed nanofluidic channels were made by balancing the etching and deposition precisely [92]. Similarly, e-beam or ion beam sculpting can reduce the pore size from initial tens of nanometers to ~1 nm, as the example shown by Fig. 9.18 in Chap. 9. These nanopores played an important role in DNA sequencing applications [93]. Pores or channels made by traditional patterning in conventional membrane materials at sub-10 nm scale are always rough in surfaces. In recent years, a class of nanomaterials such as nanotubes and 2D materials made of carbon, boron nitride, and MoS_2 are being investigated in terms of their ionic and fluid transport, because of their unique crystallographic nature and

electronic properties. Atomically flat angstrom-scale slits can be made by removing a single atomic plane from a bulk crystal [94], which provides new platforms for the development of nanofluidics [95].

References

1. Moore, G.E.: Cramming more components on integrated circuits. Electronics. **38**(8), 114 (1965)
2. Burg, D., Ausubel, J.H.: Moore's law revisited through Intel chip density. PLoS One. **16**(8), e0256245 (2021)
3. Perry, T.S.: How the father of FinFETs helped save Moore's law. In: IEEE Spectrum (2020)
4. Kamal, K.Y.: The silicon age: trends in semiconductor devices industry. J. Eng. Sci. Technol. Rev. **15**(1), 110–115 (2022)
5. Gargini, P.A.: Roadmap evolution: from NTRS to ITRS, from ITRS 2.0 to IRDS. Proc. SPIE. **10450**, 1045018 (2017)
6. The International Roadmap for Devices and Systems (IRDS): 2020 Edition. https://irds.ieee.org/ (2020)
7. Zant, P.V.: Microchip Fabrication: A Practical Guide to Semiconductor Processing, 6th edn. McGraw-Hill Education (2014)
8. Zhao, L.: All About Interconnects. Semiconductor Engineering (2017)
9. IRDS 2022 Edition: Lithography. https://irds.ieee.org/ (2022)
10. Moore, S.K.: The node is nonsense. IEEE Spectr. **57**(8), 24–30 (2020)
11. Hill, B.: Intel Details Cannonlake's Advanced 10nm FinFET Node, Claims Full Generation lead Over Rivals. Hot Hard Ware (2017)
12. Davis, S.: IBM Announces 2nm GAA-FET Technology – the Sum of "Aha!" Moments. Semiconductor Digest (2021)
13. "232-Layer NAND Flash Memory" from https://www.micron.com/ (2022)
14. Uranga, A., Verd, J., Barniol, N.: CMOS–MEMS resonators: from devices to applications. Microelectron. Eng. **132**, 58–73 (2015)
15. Wang, H., et al.: A review of system-in-package technologies: application and reliability of advanced packaging. Micromachines. **14**(6), 1149 (2023)
16. Li, T., et al.: Chiplet heterogeneous integration technology—status and challenges. Electronics. **9**(4), 670 (2020)
17. Chip-on-Wafer-on-Substrate (CoWoS) –TSMC from https://en.wikichip.org/
18. Garrou, P.: 3D IC Integration: Rumors and Ruminations. Semicopnductor International (2008)
19. Gibb, K.: 2008: The Year of Through Silicon Via? I-Micronews (2008)
20. Waldrop, M.M.: The chips are down for Moore's law. Nature. **530**, 144–147 (2016)
21. Mamaluy, D., Gao, X.: The fundamental downscaling limit of field effect transistors. Appl. Phys. Lett. **106**, 193503 (2015)
22. IRDS 2022 UPDATE: More Moore. https://irds.ieee.org/ (2022)
23. Wang, Y.: Nanosheet-based complementary transistors with a 48 nm pitch. Nat. Electron. **6**, 933–934 (2023)
24. Subramanian, S., et al.: First monolithic integration of 3D complementary FET (CFET) on 300mm wafers. In: Symposium on VLSI Technology Digest of Technical Papers - TH3.1 (2020)
25. Vandooren, A., et al.: Demonstration of 3D sequential FD-SOI on CMOS FinFET stacking featuring low temperature Si layer transfer and top tier device fabrication with tier interconnections. In: IEEE Symposium on VLSI Technology and Circuits (2022)
26. Bahaa, E.A., Saleh, M.C.T.: Fundamentals of Photonics, 3rd edn. Wiley (2020)
27. Cited from https://digital-strategy.ec.europa.eu/
28. Haley, G.: Nanoimprint Finally Finds Its Footing. Semiconductor Engineering (2023)

29. Parker, G., Charlton, M.: Photonic crystals. Phys. World. **13**(8), 29 (2000)
30. Yablonovitch, E.: Inhibited spontaneous emission in solid-state physics and electronics. Phys. Rev. Lett. **58**, 2059 (1987)
31. Murakowski, J., et al.: Fabrication of 3D photonic crystals with embedded defects. Proc. SPIE. **5347**, 181–189 (2003)
32. Biswas, U., et al.: Fabrication techniques and applications of two-dimensional photonic crystal: history and the present status. Opt. Eng. **62**(1), 010901 (2023)
33. van Xiaoren, L., et al.: Research progress in preparation and application of photonic crystals. Chin. J. Mech. Eng. **36**, 39 (2023)
34. Cai, Z., et al.: From colloidal particles to photonic crystals: advances in self-assembly and their emerging applications. Chem. Soc. Rev. **50**, 5898–5951 (2021)
35. Motokawa, R., et al.: Photonic crystals fabricated by block copolymerization-induced microphase separation. Macromolecules. **49**(16), 6041–6049 (2016)
36. Liu, S., et al.: Recent progress in responsive photonic crystals of block copolymers. J. Mater. Chem. C. **8**, 16633–16647 (2020)
37. Ball, P.: Reversing the Prism. Naturenews (2008)
38. Quan, B., et al.: Microfabrication and properties of the meta-materials. Microelectron. Eng. **83**, 1364 (2006)
39. Xia, X., et al.: Fabrication of near-infrared and optical meta-materials on insulating substrates by lift-off using PMMA/Al stack. Microelectron. Eng. **84**, 1144 (2007)
40. Quevedo-Teruel, O., et al.: Roadmap on metasurfaces. J. Opt. **21**, 073002 (2019)
41. Yoon, G., et al.: Recent progress on metasurfaces: applications and fabrication. J. Phys. D. Appl. Phys. **54**, 383002 (2021)
42. Dong Kyo, O.H., et al.: Nanoimprint lithography for high-throughput fabrication of metasurfaces. Optoelectron. **14**(2), 229–251 (2021)
43. Lee, T., et al.: Scalable and high-throughput top-down manufacturing of optical metasurfaces. Sensors. **20**, 4108 (2020)
44. Chen, Y.C., et al.: The formation and the plane indices of etched facets of wet etching patterned sapphire substrate. J. Electrochem. Soc. **159**(6), D362–D366 (2012)
45. Tsai, C.F., Su, Y.K., Lin, C.L.: Further improvement in the light output power of InGaN-based light emitting diodes by patterned sapphire substrate with KOH wet-chemical etching on sidewall. Jpn. J. Appl. Phys. **51**(1), 4 (2012)
46. Hsu, Y.P., et al.: ICP etching of sapphire substrates. Opt. Mater. **27**(6), 1171–1174 (2005)
47. Su, Y.K., et al.: Pattern-size dependence of characteristics of nitride-based LEDs grown on patterned sapphire substrates. J. Crystal Growth. **311**, 2973–2976 (2009)
48. Zhou, Q., et al.: Improved efficiency of GaN-based green LED by a nano-micro complex patterned sapphire substrate. IEEE Photon. Technol. Lett. **29**(12), 983–986 (2017)
49. Oh, T.S., et al.: GaN-based light-emitting diodes on micro-lens patterned sapphire substrate. Jpn. J. Appl. Phys. **47**(7), 5333–5336 (2008)
50. Byeon, K.J., et al.: Fabrication of SiNx-based photonic crystals on GaN-based LED devices with patterned sapphire substrate by nanoimprint lithography. Opt. Express. **20**(10), 11423–11432 (2012)
51. Gao, H., et al.: Fabrication of nano-patterned sapphire substrates and their application to the improvement of the performance of GaN-based LEDs. J. Phys. D. Appl. Phys. **41**, 115106 (2008)
52. Lee, Y.-T.H.-C.: Metal contact printing photolithography for fabricating sub-micrometer patterned sapphire substrates in light-emitting diodes. In: Proceedings of IEEE-NEMS, Tokyo (2012)
53. Zhou, S., et al.: Application of patterned sapphire substrate for III-nitride light-emitting diodes. Nanoscale. **14**, 4887 (2022)
54. National Nanotechnology Initiative from https://www.nano.gov/
55. NNI Supplement to the President's 2023 Budget from https://www.nano.gov/
56. BCC Research Team: Global Nanotechnology Market. BCC Publishing (2023)

57. Mikolajick, T., Weber, W.M.: Chapter 1 silicon nanowires: fabrication and applications. In: Li, Q. (ed.) Anisotropic Nanomaterials: Preparation, Properties and Applications. Springer (2015)
58. Cao, A., Sudhölter, E.J.R., de Smet, L.C.P.M.: Silicon nanowire-based devices for gas-phase sensing. Sensors. **14**, 245–271 (2014)
59. Weber, W.: Silicon nanowires: catalytic growth and electrical characterization. Phys. Status Solidi B. **243**, 3340 (2006)
60. Merkulov, V.I., et al.: Patterned growth of individual and multiple vertically aligned carbon nanofibers. Appl. Phys. Lett. **76**(24), 3555 (2000)
61. Teo, K.B.K., et al.: Uniform patterned growth of carbon nanotubes without surface carbon. Appl. Phys. Lett. **79**(10), 1534 (2001)
62. Robertson, J.: Growth of nanotubes for electronics. Materialstoday. **10**(1–2), 36 (2007)
63. Koblischka, M., Koblischka-Veneva, A.: Fabrication of superconducting nanowires using the template method. Nano. **11**(8), 1970 (2021)
64. Wang, W.L., Meng, S., Kaxiras, E.: Graphene NanoFlakes with large spin. Nano Lett. **8**(1), 241–245 (2008)
65. George, A., Mathew, S., van Gastel, R.: Large area resist-free soft lithographic patterning of graphene. Small. **9**(5), 711–715 (2013)
66. Shi, Z., Yang, R., Zhang, L.: Patterning graphene with zigzag edges by self-aligned anisotropic etching. Adv. Mater. **23**, 3061–3065 (2011)
67. Rahman, M., et al.: A review on nanomaterial-based additive manufacturing: dynamics in properties, prospects, and challenges.. Progress in Additive Manufacturing 2023 (published online: https://doi.org/10.1007/s40964-023-00514-8)
68. Cui, Z., et al.: Printed Electronics: Materials, Technologies and Applications. Wiley (2016)
69. Corzo, D., et al.: Flexible electronics: status, challenges and opportunities. Front. Electron. **1**, 594003 (2020)
70. Shin, S.J., et al.: Room-temperature charge stability modulated by quantum effects in a nanoscale Silicon Island. Nano Lett. **11**(4), 1591–1597 (2011)
71. Seabaugh, A.: The tunneling transistor. In: IEEE Spectrum (2013)
72. Collaert, N., et al.: Ultimate nano-electronics: new materials and device concepts for scaling nano-electronics beyond the Si roadmap. Microelectron. Eng. **132**, 218–225 (2015)
73. Chen, Y., et al.: Sub-10 nm fabrication: methods and applications. Int. J. Extrem. Manuf. **3**, 032002 (2021)
74. Kanungo, S., et al.: 2D materials-based nanoscale tunneling field effect transistors: current developments and future prospects. npj 2D Mater. Appl. **6**, 83 (2022)
75. A. F. Koenderink; et al, Nanophotonics: Shrinking Light-Based Technology. Science, 2015. 348(6234): p. 516–521
76. Lin, L., Zheng, Y.: Optimizing plasmonic nanoantennas via coordinated multiple coupling. Sci. Rep. **5**, 14788 (2015)
77. Huck, C., et al.: Surface-enhanced infrared spectroscopy using nanometer-sized gaps. ACS Nano. **8**, 4908–4914 (2014)
78. Yoon, S.J., et al.: Hopping of single nanoparticles trapped in a plasmonic double-well potential. Nano. **9**(16), 4729–4735 (2020)
79. Cui, A., et al.: Single grain boundary break junction for suspended Nanogap electrodes with Gapwidth down to 1–2 nm by focused ion beam milling. Adv. Mater. **27**, 3002–3006 (2015)
80. Chengfeng, P., et al.: Nanofabrication: the unsung hero in enabling advances in nanophotonics. Nano. **12**(8), 1359–1361 (2023)
81. Sivakami, A., Sarankumar, R., Vinodha, S.: Introduction to nanobiotechnology: novel and smart applications. In: Pal, K. (ed.) Bio-Manufactured Nanomaterials. Springer (2021)
82. Cleland, A.N., Roukes, M.L.: Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystal. Appl. Phys. Lett. **69**(8), 2653 (1996)
83. Lang, H.P., Hegner, M., Gerber, C.: Cantilever array sensors. Mater. Today. **8**(4), 30–36 (2005)
84. Ilic, B., Yang, Y., Aubin, K.: Enumeration of DNA molecules bound to a Nanomechanical oscillator. Nano Lett. **5**(5), 925–929 (2005)

85. Yang, Y.T., et al.: Surface Adsorbate fluctuations and noise in NanoelectromechanicalSystems. Nano Lett. **11**, 1753–1759 (2011)
86. Xu, B., et al.: Nanomechanical resonators: toward atomic scale. ACS Nano. **16**(10), 15545–15585 (2022)
87. Terry, S.C., Jerman, J.H., Angell, J.B.: A gas chromatographic air analyzer fabricated on silicon wafer. IEEE Trans. Electron. Devices. **26**, 1880 (1979)
88. Hajam, M.I., Khan, M.M.: Microfluidics: a concise review of the history, principles, design, applications, and future outlook. Biomater. Sci. **12**, 218–251 (2024)
89. Bocquet, L., Charlaix, E.: Nanofluidics, from bulk to interfaces. Chem. Soc. Rev. **39**, 1073–1095 (2010)
90. Haywood, D.G., et al.: Fundamental studies of Nanofluidics: Nanopores, Nanochannels, and Nanopipets. Anal. Chem. **87**, 172–187 (2015)
91. Duan, C., Wang, W., Xie, Q.: Review article: fabrication of nanofluidic devices. Biomicrofluidics. **7**(2), 026501 (2013)
92. Nam, S.-W., Lee, M.-H., Lee, S.-H.: Sub-10-nm Nanochannels by self-sealing and self-limiting atomic layer deposition. Nano Lett. **10**, 3324–3329 (2010)
93. MacKenzie, M., Argyropoulos, C.: An introduction to Nanopore sequencing: past, present, and future considerations. Micromachines. **14**, 459 (2023)
94. Esfandiar, A., et al.: Size effect in ion transport through angstrom-scale slits. Science. **358**, 511–513 (2017)
95. Bocquet, L.: Nanofluidics coming of age. Nat. Mater. **19**, 254–256 (2020)

Index

A
Aberrations, 84, 85, 91–95, 97, 98, 106, 130, 144
Absorbance modulation layer (AML), 342
Absorbers, 25, 27, 42, 315
Acetone, 16, 205, 220, 242, 306
Adhesion lithography, 330, 331
Aerial images, 39, 41, 42, 46, 50, 53, 65
Aerosol jet printing, 322
Alignment, 15, 96, 173, 205, 263, 319, 370
Alignment marks, 15, 97, 214, 230, 249
Alternating PSMs, 39, 41, 42
Amorphous material, 150
Anisotropic, 259–267, 269, 271, 279, 281, 283, 334, 336
Annealing, 210, 289, 370–373, 391
Anodized aluminum oxide (AAO), 215, 248, 373, 376, 397
Antireflection coating (ARC), 16, 37, 224
Apertures, 12, 35, 37, 38, 61, 63, 70, 86, 91, 93, 96, 130, 133–135, 145, 147, 172, 173, 178, 179, 285, 319, 320
Aspect ratios, 125, 212, 265, 267, 276, 282, 284, 285, 306, 314, 317, 376
Astigmatism, 91, 94, 96
Atomic force microscope (AFM), 52, 169, 171–173, 182–189, 192, 193, 265, 396
Atomic layer deposition (ALD), 181, 290, 299–302, 317, 330, 339, 388, 402
Atomic layer etching (ALE), 289–290, 388
Attenuated PSM (Half-tone PSM), 38, 39, 41, 42

B
Backscatterings, 105, 106, 113, 121, 123, 130, 131, 163
Beam blanking, 96, 102, 134
Beam pen lithography, 190
Black silicon, 288
Block copolymer lithography, 368–373
Block copolymers (BCPs), 7, 368–373, 387, 393
Boiling temperatures, 271, 272
Bosch process, 280–283, 287
Bottom-up, 5, 349, 350, 361, 393, 396, 398
Bowtie, 63, 173, 179, 194
Bright field, 17
Brownian motion, 358
Bulk micromachining, 265

C
Capillary forces, 54, 55, 221, 223, 226, 228, 236, 241–245, 267, 351, 358, 359, 362, 365
Carbon nanotubes (CNTs), 267, 322, 324, 349, 352, 366, 396–398
Casting, 228, 235, 367
Channeling effect, 150–152, 288
Chemically amplified resists (CARs), 17, 19, 26, 50, 121, 129, 175, 194, 391
Chemical mechanical polishing (CMP), 30, 58, 317
Chemical vapor depositions (CVDs), 180, 299, 300, 302, 317, 329, 351, 386

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Z. Cui, *Nanofabrication*, https://doi.org/10.1007/978-3-031-62546-6

Chemoepitaxy, 370, 371, 373
Chiplets, 390
Chips, 3, 9, 26, 27, 30, 35, 71, 156–158, 195, 201, 202, 279, 324, 384–388, 390, 391, 402
Chromatic aberration, 91–94, 118, 130, 144
Chromeless phase lithography (CPL), 42, 43
Chromeless PSM, 39
Clear defects, 158–160
Colloidal crystals, 358–361, 364–366
Colloidal particles, 349, 358–363, 365, 366
Complementary FET (CFET), 390, 391
Computational lithography, 47, 48
Contact printing, 59, 63, 69, 239, 241, 397
Contrasts, 2, 18, 19, 29, 31, 37, 42, 49–52, 54, 65, 94, 113, 119, 120, 124–126, 129, 155, 159, 162, 164, 229, 240, 314, 341, 363, 394
Critical dimension (CD), 19, 26, 30, 45, 50, 52–54, 181
Cross-over, 89, 90, 93, 134, 366
Cryogenic processes, 283

D
Damascene, 317
Dark field, 17, 68
Deep reactive ion etching (DRIE), 279–289, 317, 388
Deep structures, 278, 281, 283
Deep UV lithography (DUV), 9, 11, 17, 21, 22, 26, 30, 38, 42, 50, 226, 228, 387
Demolding, 203, 205, 211–213, 218, 220, 223, 225, 233, 234
Depth of focus (DOF), 29–31, 37
Descum, 16, 261
Design for manufacturability (DFM), 58
Design rules, 58
Design-technology co-optimization (DTCO), 57–58
Developers, 16, 17, 20, 34, 120, 121, 127–129, 197, 220, 308, 312
Development, 9, 85, 142, 169, 201, 261, 308, 327, 350, 383
Die, 3, 33, 228, 249, 279, 389, 390
Dip-pen nanolithography (DPN), 188–190, 196, 197, 260
Directed self-assembly (DSA), 371–373, 387
Displacement Talbot lithography (DTL), 65
DNA, 187, 201, 349, 366, 401, 402
DNA origami, 366
Dry etch, 221, 223, 224
Dry etching, 224, 257, 258, 267

E
Edge lithography, 60
Elastic scatterings, 103, 148
Electrochemical deposition, 192
Electroforming, 314
Electrohydrodynamic jet (E-jet), 322, 323
Electroless plating, 314
Electron beam lithography, 7, 25, 46, 83–135, 147, 151, 156, 162, 201, 265, 306, 308, 311, 344, 371, 393, 396
Electron beam resists, 162
Electron column, 95, 96
Electron gun, 91, 93, 96, 291
Electron lens, 85–89, 96, 130, 175
Electron projection lithography (EPL), 4, 132, 164
Electron proximity effect, 97, 108
Electrophoretic deposition, 364
Electroplating, 205, 300, 314–317
End point detection, 155, 159
Etching, 2, 16, 118, 144, 177, 204, 257, 299, 327, 366, 386
Etch selectivity, 19, 26, 177, 180, 221, 223, 225, 273, 279, 291, 372
Etch stop, 286
Evanescent near-field optical lithography (ENFOL), 59, 60
Evanescent wave, 59, 60, 62
Evaporation induced self-assembly (EISA), 356, 357
Excimer lasers, 9, 10, 21, 49, 67, 72
Exposure doses, 15, 18–20, 24, 53, 74, 91, 93, 94, 97, 108–115, 118, 120–128, 162–164, 176, 177, 223, 308, 309, 336, 343, 344
Exposure latitude, 19
Extreme UV (EUV), 7, 9, 10, 22–26, 32, 34, 42, 50, 53, 55, 57, 58, 66, 69, 83, 84, 122, 132, 134, 135, 160, 161, 165, 169, 201, 227, 330, 373, 387

F
Fast secondary electrons, 114
Femtosecond lasers, 72, 74, 342
Fidelity, 44, 207, 217, 221
Field electron emission, 143, 175
Field emission cathodes, 90, 91, 96, 174
Field evaporation, 143, 191, 192
Field ion microscope (FIM), 146
Field stitching, 98, 99, 102, 228
FinFETs, 158, 384, 385, 388, 390, 391
Flatness, 25, 27, 30, 228

Focused ion beam (FIB), 5, 7, 25, 141–165, 179, 247, 248, 257, 288, 291, 299, 319, 327, 335, 337, 339, 340, 344–346, 400–402
Forward scatterings, 104–106, 109, 113, 117, 130, 131, 163
Friction-induced patterning, 185

G
GAA FET, 384, 385, 388, 390
Gallium staining, 26, 144, 160
Gas field ion source (GFIS), 141, 142, 146–148, 155, 161, 163, 164
Gaussian beam, 98, 101, 104, 132, 133
Glass transition temperature, 20, 180, 203, 207, 211, 215, 234, 300
Graphoepitaxy, 371, 373
Gray scale lithography, 227
Grayscale mask, 45

H
Half-pitch (hp), 25, 36, 57, 66, 131, 165, 177, 330, 332, 385, 387, 388
Heat-assisted magnetic recording (HAMR), 202
Hexadecanethiol (HDT), 240, 241, 353, 354
High aspect ratios, 19, 27, 54, 124, 125, 130, 185, 192, 212, 213, 216, 217, 221, 222, 225, 229, 233, 234, 237, 267, 279, 282, 299, 300, 313–315, 317, 322, 374, 388, 391
Holographic lithography, 66
Hot embossing, 5, 203
Hot spots, 58
Hydrogen bonds, 351, 355
Hydrogen depassivation lithography (HDL), 181, 182, 197
Hydrogen silsesquioxane (HSQ), 120–122, 124–126, 129–131, 164, 165, 206, 207, 215–217, 293, 311–313, 335, 344
Hydrophilic, 190, 245, 363
Hydrophilicity, 363
Hydrophobic, 211, 219, 244, 352, 363
Hydrophobicity, 363

I
Immersion lithography, 9, 32–34, 55, 66, 132, 227, 330
Imprinting proximity effect, 209
Imprints, 202, 203, 215, 226, 230, 232–235, 238, 245–247, 250, 375
Inductively coupled plasma (ICP), 145, 147, 155, 218, 276–279, 282, 289
Inelastic scatterings, 103, 104, 148, 176
Inkjet printing, 321–323, 327, 359
Intensity distributions, 11, 12, 36, 38, 39, 47, 65, 67, 342, 343
Interferometric lithography (IL), 10, 65–69
Intermediate polymer stamp (IPS), 219
International Roadmap for Devices and Systems (IRDS), 373, 385, 386, 390, 399
International Technology Roadmap for Semiconductors (ITRSs), 36, 52, 201, 371, 385, 386
Inverse lithography technology (ILT), 47, 48, 57, 134, 135
Ion beam milling, 248, 291, 292, 401, 402
Ion beam sculpting, 339, 346, 402
Ion bombardments, 150, 154, 260, 270, 274, 276, 279, 280, 289, 290
Ion implantation, 141, 142, 144, 150, 156, 291, 386
Ion induced damage, 284, 288–289
Ion projection lithography (IPL), 4, 164, 165
Isotropic, 57, 185, 259–265, 269, 270, 274, 280, 288, 290, 299, 305, 329, 332
Isotropic etch, 57, 259–261, 265, 269, 274, 280, 329, 332

J
Jet and flash imprint lithography (J-FIL), 225–228, 230, 232, 242

K
Knudsen's cosine law, 302, 305

L
Lab-on-a-chip, 402
Langmuir–Blodgett (LB) films, 352
Laser direct write, 46, 69, 84, 133, 248, 344
Lateral addition, 337, 339
Lateral subtraction, 332, 333, 335
Lattice-like, 5
Lift-off, 122, 131, 177, 178, 180, 186
Lift-off resist (LOR), 310, 311
LIGA, 27, 314, 315
Light emitting diode (LED), 394, 395, 398
Line edge roughness (LER), 26, 49, 52–55, 60–63, 122, 128, 227, 387
Line width roughness (LWR), 49, 52–55

Liquid metal alloy ion source (LMAIS), 144
Liquid metal ion source (LMIS), 141–148, 155, 163, 164, 344
Litho-etch-litho-etch (LELE), 55, 56
Litho-freeze-litho-etch (LFLE), 55, 56
Loading, 283–285
Local oxidation of silicon (LOCOS), 334
Long range ordered, 351, 369
Low-pressure CVD (LPCVD), 260, 300, 301

M
Mask aligner, 10, 15, 59, 63, 65, 72, 214, 230, 343
Mask defects, 158–160
Mask error enhancement factor (MEEF), 53, 54
Maskless lithography (ML2), 69–71, 134
Masters, 5, 6, 203, 205, 214, 215, 219, 235, 236
Mercury lamps, 9, 10, 21, 49, 75
Metal assisted chemical etching (MACE), 265–267
Metal-ligand, 355
Metal-organic frameworks (MOFs), 357
Metamaterials, 63, 181, 393, 394
Metasurfaces, 394
Micro-aspiration assisted lithography (MAAL), 242, 244
Microcontact printing (μCP), 207, 236, 238–241, 260, 353, 363
Microdomains, 368–370
Microelectromechanical systems (MEMSs), 59, 70, 72, 279, 281, 314, 388
Microlens, 184, 249
Micromolding in capillaries (MIMIC), 236, 241
Microoptical elements, 394
Microphase separation, 368–372
Microstereolithography, 72, 73
Microtrenching, 283, 285–286
Micro zone plate (MZP), 315
Minimum feature size, 6, 22, 98, 385, 390
Mobility, 274, 288, 289, 340, 346, 352, 353, 355, 370, 399
Moiré fringes, 231, 232
Molecular beam epitaxy (MBE), 299, 332
Molecular self-assembly, 7, 349–352, 358
Monte Carlo method, 104, 149, 150
Moore's law, 384, 385, 388, 390
Multi-beam mask writer (MBMW), 132, 134, 135, 165
Multilayer resist, 55, 62, 118, 177, 216

N
Nanobiotechnology, 401–403
Nanoelectronics, 366, 396, 399–400
Nanofluidics, 402, 403
Nanoimprint, 5, 201, 203–215, 221, 226, 232–236, 238, 245–250, 299, 327, 329, 332, 375, 376, 394, 395
Nanoimprint lithography (NIL), 201–250
Nanoparticles, 122, 202, 324, 356, 361–366, 376, 396–398
Nanophotonics, 396, 400–401
Nanopores, 337, 339, 340, 346, 374–377, 402
Nanosphere, 366, 368
Nanosphere lithography, 366–368, 394
Nanotechnologies, 7, 84, 197, 202, 226, 228, 232, 383, 395–403
Near-field optical lithography, 10, 59–63, 193
Near-field scanning optical microscope (NSOM), 172, 173, 177–179, 196
Negative resist, 18, 19, 49, 119, 120, 311
Negative tone development (NTD), 17, 18
Next generation lithography (NGL), 32, 132, 164, 201
Non-covalent bond, 355
Notching, 284, 286–287
Numerical aperture (NA), 9, 13, 14, 25, 28–37, 45, 104, 169, 315, 387

O
Off-axis illumination (OAI), 37–38, 43, 62
Opaque defects, 25, 158–160
Optical lithography, 1, 9, 83, 156, 177, 201, 306, 337, 367, 383
Optical proximity correction (OPC), 37, 41, 44–48, 55, 57, 58, 66, 69, 227, 260
Optical proximity effects, 44, 45, 108, 227
Optical steppers, 27, 71, 215
Organic light emitting diode (OLED), 249, 321
Oven, 15, 16, 216
Overcut, 308, 311
Overlay, 15, 26, 98, 130, 202, 214, 232, 249, 250
Oxidation sharpening, 333

P
Passivation layer, 181, 276, 354
Patterned sapphire substrate (PSS), 394, 395
Pattern generator, 71, 97, 101, 164

Pattern transfer, 1, 2, 4–7, 16, 19, 27, 62, 118, 162, 177, 194, 206, 207, 209–211, 217, 221, 224–226, 241, 257–294, 299–324, 327, 328, 337, 349, 354, 372, 383, 393
Pellicle, 22
Phase-shift masks (PSMs), 37–41, 47, 57, 260
Photoacid, 26, 49, 52, 125, 210
Photoinitiator, 223
Photomasks, 9, 15, 22, 37, 38, 41, 47, 61, 62, 68, 69, 72, 84, 98, 117, 132, 135, 160, 248, 260, 261, 327
Photonic bandgap, 392, 393
Photonic crystals, 68, 74, 202, 361, 366, 392, 393
Photoresist, 9, 83, 160, 177, 217, 274, 308, 329, 362, 386
Physical vapor deposition (PVD), 299, 300, 304, 351
Planar technology, 258, 385
Plasma, 6, 16, 218, 261, 299, 332, 366, 397
Plasma enhanced CVD (PECVD), 300, 301, 397
Plasma etches, 271
Plasma ion sources, 141, 142, 145–147, 155
Point spread functions (PSFs), 104–107, 114, 115, 123, 128, 163
Polarization, 31, 66
Polycrystalline material, 150, 340
Poly-dimethyl-siloxane (PDMS), 63, 196, 228, 229, 235–245, 249
Polymerization, 72, 73, 75, 223, 342
Polymer pen lithography (PPL), 196
Poly methyl methacrylate (PMMA), 62, 83, 84, 104, 108, 114, 117, 121–123, 125–131, 163, 175, 177, 178, 180, 185–188, 194, 201, 205–210, 212, 213, 216, 217, 233–235, 241–243, 265, 293, 306, 308–313, 342, 344, 372
Polyphthalaldehyde (PPA), 180, 187, 195
Polysilicon, 269, 329
Polystyrene (PS), 209, 233, 243, 267, 358, 359, 362, 364–367, 370, 372
Polystyrene-block-poly-poly (methyl methacrylate) (PS-b-PMMA), 372
Porous alumina, 373–377
Positive resists, 17–19, 108, 120, 293, 311
Post exposure bake (PEB), 15, 33, 49, 50, 53
Post-exposure delay, 126
Post-trimming, 335
Power, Performance, Area, Cost (PPAC), 390
Prebake (soft bake), 15, 216, 232
Precursor gas, 154, 156, 160, 192, 289, 290
Printed electronics, 249, 321, 322, 398
Process latitude, 20
Profiles, 50, 51, 58, 75, 108–110, 112–114, 123–125, 184–186, 211, 212, 216, 217, 224, 234, 259, 261, 262, 265, 271, 276, 281, 284, 293, 294, 307–313, 336, 344–346
Projection lithography, 3–5, 10–14, 63, 68, 132
Projection Mask-Less Patterning (PMLP), 165
Proximity effect correction (PEC), 111, 112, 131
Proximity printing, 344

R

Raster scan, 69, 75, 98, 99, 102, 113, 171
Reactive ion etch (RIE), 205, 207–209, 211, 216–218, 220, 224–226, 237, 260, 267–279, 285, 286, 288–293, 317, 328, 329, 332, 333, 335–337, 339
Recoiling atoms, 149, 151, 152
Redeposition, 144, 153, 154, 270, 294
Reduced graphene oxide (rGO), 188
Refractive index, 13, 26, 31, 32, 34, 66, 85, 91, 392, 394
Release agent, 211, 217
Resist, 1, 17, 83, 169, 201, 257, 299, 335, 354
Resist stripping, 311
Resolution enhancement techniques (RETs), 9, 37, 38, 55, 57, 58, 68–70, 84, 260
Reticles, 9, 70
Rolling mask, 63
Roll-to-roll (R2R), 63, 206, 245–250, 394

S

Sacrificial layer, 391
Scanning electron microscope (SEM), 47, 50, 52, 53, 83, 93, 94, 97, 101, 102, 119, 123, 130, 147, 157, 174, 179, 193, 288, 312, 316, 317, 319, 320, 331, 332, 339, 343, 345, 346, 362, 364
Scanning probe lithography (SPL), 169, 180, 185, 190, 192–194, 197, 371
Scanning probe microscope (SPM), 170, 193
Scanning transmission electron microscope (STEM), 94, 131, 161–162
Scanning tunneling microscope (STM), 169–173, 175–178, 180, 182, 183, 191–194, 197, 396
Scattering bars, 45
Scattering cross-sections, 103, 148, 149
Schottky cathode, 91
Screen printing, 321

Seamless roller, 248, 249
Self-aligned double patterning (SADP), 55–57, 330
Self-amplified depolymerization, 180
Self-assembled monolayer (SAM), 187–189, 238, 240, 241, 260, 330, 352–354, 356, 363
Self-assembly, 5, 350, 387
Self-construction, 349
Sensitivity, 18, 19, 26, 49–50, 52, 53, 90, 117–126, 129, 130, 132, 162–164, 194, 223, 308, 344
Shadow masks, 299, 318, 321, 322
Shadow sphere lithography, 366
Shaped beam, 133, 134
Shear stress, 212, 234
Sidewall lithography, 328–332, 340
Sidewalls, 124, 154, 173, 211, 212, 218, 240, 262, 263, 265, 276, 280–283, 285, 286, 294, 305, 328–330, 336, 337
Silicon on insulator (SOI), 195, 286, 287, 336, 397
Silicon wafers, 3, 4, 15, 16, 27, 30, 159, 205, 214, 215, 249, 260–265, 279, 281, 286, 321, 341, 402
Soft lithography, 7, 203, 235–245, 327
Soft stamps, 207, 228, 236–238, 353
Solvent-assisted microcontact molding (SAMIM), 236, 242, 243
Space charge effect, 85, 93, 94, 118, 130
Spatial light modulator (SLM), 71, 72
Spherical aberration, 91–93, 96, 130
Spin-coating, 20, 216, 223, 232–234, 249, 321
Spray coating, 15, 299
Sputtering, 5, 141, 267, 299, 345, 366, 386
Sputtering yields, 150–154, 291, 307
Stamp, 5, 202, 315
Step and flash imprint lithography (SFIL), 225, 226, 231
Step & stamp imprinting lithography (SSIL), 214, 226
Subdiffraction, 342
Sub-resolution assist features (SRAFs), 45, 46, 55, 57, 69
Substrate conformal imprint lithography (SCIL), 228–232, 237, 242
Superparamagentic colloidal particles, 365
Super-resolution, 342–346
Supporting membrane, 27
Surface energy, 151, 190, 211, 223, 229, 233, 237, 363
Surface micromachining, 70
Surface plasmon polarization (SPP), 60–62, 194
Surface tension, 33, 55, 233, 351, 358–360, 365
Surfactants, 353, 356, 363
Synchrotron radiation source (SRS), 23, 27
System in package (SiP), 388, 389
System-on-chip (SoC), 279, 317, 388

T
Talbot lithography, 63–65
Tapping mode, 171, 172, 182, 183
Technology nodes, 10, 35, 36, 48, 52, 57, 126, 373, 385, 388
Templates, 6, 202, 215, 226–228, 231, 350, 361, 363, 365, 366, 376, 397
Thermal evaporation, 300, 302, 303, 305, 307, 308, 318, 319, 321, 366, 386
Thermal nanoimprint lithography (T-NIL), 203, 205, 207, 209, 210, 214, 215, 217–219, 223, 226, 232, 247
Thermal oxidation, 329, 333, 334, 341, 344
Thermal scanning probe lithography (t-SPL), 180, 197
Thermionic cathode, 89, 90, 93, 291
Thermomechanical indentation, 180, 187, 188
Thick photoresist, 314, 393
Thin film deposition, 299–308, 339, 351, 385
Thin film transistors (TFTs), 398
3D integration, 388
Through-silicon-via (TSV), 279, 317, 389, 390
Tip-based nanofabrication (TBN), 169–197
Topcoats, 33, 34
Top-down, 5, 6, 349, 350, 361–363, 371, 376, 393, 396, 397
Topography, 30, 150, 162, 186, 308, 361–363, 371
Top surface imaging (TSI), 37, 185
Tribo-nanolithography, 185
Tunnel field-effect transistor (TFET), 399
Two-color photolithography, 342
Two-photon polymerization (2PP), 72–75

U
Ultrasonic assisted development, 127, 129
Undercut, 212, 216, 308–313
UV cured nanoimprint lithography (UV-NIL), 202, 217–221, 223, 225, 226, 228–231, 238

V
Van der Waals forces, 351, 353, 355, 358, 362
Variable shaped beam (VSB), 101, 132–134
Vector scan, 97–99, 101, 102, 113
Vertical subtraction, 336
Via holes, 314, 317, 386
Virtual source, 89, 90, 92
Volume production, 21, 34, 72, 321

W
Wafer stage, 33, 230
Wet etch, 236, 263, 299
Wettability, 363
Work function, 174

X
X-ray lithography, 22, 27, 119, 132, 315, 343

Y
Young's single slit diffraction, 11